

Titre: Optimisation du positionnement des annonces textuelles en
Title: marketing interactif

Auteur: Koukla Azeuli Nkamegni
Author:

Date: 2017

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Azeuli Nkamegni, K. (2017). Optimisation du positionnement des annonces
Citation: textuelles en marketing interactif [Thèse de doctorat, École Polytechnique de
Montréal]. PolyPublie. <https://publications.polymtl.ca/2906/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/2906/>
PolyPublie URL:

**Directeurs de
recherche:** Michel Gamache, & Alain Hertz
Advisors:

Programme: Doctorat en mathématiques de l'ingénieur
Program:

UNIVERSITÉ DE MONTRÉAL

OPTIMISATION DU POSITIONNEMENT DES ANNONCES TEXTUELLES EN
MARKETING INTERACTIF

KOUKLA AZEULI NKAMEGNI
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(MATHÉMATIQUES DE L'INGÉNIEUR)
DÉCEMBRE 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

OPTIMISATION DU POSITIONNEMENT DES ANNONCES TEXTUELLES EN
MARKETING INTERACTIF

présentée par : AZEULI NKAMEGNI Koukla
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. ADJENGUE Luc-Désiré, Ph. D., président
M. GAMACHE Michel, Ph. D., membre et directeur de recherche
M. HERTZ Alain, Doctorat ès Sc., membre et codirecteur de recherche
M. CAPOROSSİ Gilles, Ph. D., membre
Mme REKİK Monia, Ph. D., membre externe

DÉDICACE

*On dit que ça prend tout un village pour élever un enfant.
Je dédie cette thèse au village qui m'a élevée et continue de le faire
tout particulièrement à mes très chers parents
et à ma grand-mère Rose partie bien trop tôt et qui restera pour moi un exemple de
dévouement et de pondération.*

REMERCIEMENTS

Tout d'abord, je tiens à remercier mes directeur et codirecteur Michel Gamache et Alain Hertz pour leur gentillesse, leur patience et leur disponibilité tout au long de ces cinq dernières années.

Merci aussi à Sandrine Paroz, pour sa disponibilité, son aide et ses conseils surtout au début du projet.

Je voudrais également remercier chaleureusement tous mes amis et collègues du programme de tutorat au SEP pour les fous rires et autres bons moments.

Un remerciement spécial à ma famille et ma famille de cœur pour les encouragements et le soutien depuis mon premier jour à Poly.

Pour finir, il n'y a pas de mots pour exprimer ma gratitude envers mes parents qui, même lorsqu'ils n'approuvaient pas mes choix, n'ont jamais cessé de croire en moi et se sont toujours assurés que je sache que j'avais tout leur soutien. Ça fait plus de 27 ans que ça dure et vous n'avez jamais failli. Merci !

RÉSUMÉ

Lorsqu'un utilisateur tape du texte dans le champ réservé à cet effet dans un moteur de recherche, pendant le chargement de la page des résultats, une enchère a lieu pour déterminer quelles annonces textuelles seront affichées et dans quel ordre. Le texte tapé par l'utilisateur doit pouvoir être associé à au moins un des mots-clés choisis par un annonceur pour que ce dernier soit éligible à participer à cette enchère.

Plus une annonce est affichée haut sur la page, mieux elle est classée et ; mieux elle est classée, plus l'annonceur devra payer pour chaque clic sur ladite annonce et plus les utilisateurs cliqueront dessus. Toutefois, le montant facturé par le moteur de recherche à un annonceur pour chaque clic ne dépasse jamais l'enchère de cet annonceur. De plus, l'annonceur peut fixer un budget périodique (par jour, semaine ou mois) que le moteur de recherche devra respecter. Les annonceurs doivent donc établir une stratégie d'enchère qui leur permettra de maximiser la performance de leurs campagnes tout en respectant leur budget. Malheureusement, ils n'ont accès qu'à une quantité limitée d'informations pour le faire.

Dans le cadre de ce projet de recherche, un modèle d'optimisation est proposé. Il permet de déterminer la meilleure position pour chaque mot-clé avec pour objectif d'optimiser les profits générés par une ou plusieurs campagnes d'un annonceur ou d'autres indices de performance de campagne tels que les revenus ou les conversions. Ce modèle, présenté dans le troisième chapitre, est basé sur les données de navigation partielle des utilisateurs. Elles sont représentées par un graphe et le problème est formulé comme un problème de flot ou de flot à coût minimum selon l'objectif utilisé. Le modèle n'étant pas linéaire, nous avons également développé cinq algorithmes pour résoudre ce problème.

Quatre d'entre eux sont basés sur l'algorithme tabou. Dans la version élémentaire de l'algorithme, on modifie d'une unité la position d'un mot-clé et on recalcule le flot en utilisant des formules liant la position d'un mot-clé au nombre de clics espérés sur celui-ci ainsi qu'au coût-par-clic moyen de ce mot-clé. Le mot-clé dont la position est changée est choisi de sorte que la modification soit la plus profitable possible tout en respectant les contraintes de budget. Par la suite, il devient tabou de modifier à nouveau la position du même mot-clé pendant un nombre prédéterminé d'itérations.

Les trois autres algorithmes sont des variantes du premier. Dans la première variante, la position peut être modifiée de plus d'une unité. Dans la seconde, on accepte avec une pénalité que le budget soit dépassé. Finalement, dans la troisième version, les variantes des deux premières versions sont implémentées.

Le cinquième algorithme est basé sur l'algorithme glouton. À chaque itération, pour chaque

mot-clé, on détermine s'il est préférable de maintenir sa position ou de la modifier d'une unité. Si changer de position est plus profitable, alors on détermine quelle position est la plus profitable et on effectue le changement. Ce changement peut être de plus d'une unité. Les dépassements de budget sont permis mais pénalisés. Cet algorithme s'est avéré être le plus rapide avec des résultats comparables à ceux obtenus avec les tabous.

Dans le quatrième chapitre, on propose une étude comparative sur l'impact de différentes modifications, celles-ci étant apportées soit au modèle soit à un paramètre. La première étude consiste à modifier la fonction-objectif et de mesurer l'impact de ce changement sur les résultats. Nous avons comparé les valeurs de différentes mesures de performances lorsque l'objectif est d'optimiser une d'elles à la fois. L'analyse révèle que le profit et les coûts sont les plus sensibles au changement et que la maximisation des visites est l'objectif ayant l'impact le plus négatif sur les autres mesures de performance.

Une seconde étude analyse l'impact de la variation du budget sur les résultats de la campagne. Dans cette étude, on constate que la maximisation du profit est peu sensible à la variation de budget et que, pour les autres objectifs, une variation de $\pm 10\%$ ou 20% ne se traduit pas nécessairement en une variation de performance de $\pm 10\%$ ou 20% .

Finalement, la dernière modification analysée est la variation du coût-par-clic sur l'ensemble des mots-clés. Pour cette analyse, les coûts-par-clic de tous les mots-clés sont aléatoirement modifiés afin de voir si ces changements ont un impact significatif sur les valeurs des indices de performance et sur les positions des mots-clés. Dans le cas de la maximisation des profits, moins de 7% des mots-clés se sont retrouvé à des positions différentes au final et la différence était seulement d'une unité. Pour le reste des objectifs, les variations sont plus importantes.

ABSTRACT

When a user types a text in the search field of a search engine, during the loading of the page, an auction takes place to determine which text ads will appear along with the search results and in which order. The text typed by the user must be associated with at least one of the keywords chosen by an advertiser for that advertiser to be eligible to participate in that auction. In this thesis, we only consider the first ten positions which are all on the first page as all the others have very little impact on the ad performance.

So, the highest an ad is placed on the result page, the better it is ranked, the more each click will cost and the more clicks it will receive. However, the cost-per-click billed to the advertiser never exceeds his bid (the maximal amount he is willing to pay per click). Moreover, the advertiser can set a periodical budget. A period might be a day, a week or a month. That budget is to be respected by the search engine. The advertiser must find the best bidding strategy that will help them maximize the performance of their ad campaigns while respecting their budget.

Thus, in the third chapter, a new model to determine which position to aim for on search engines using partial users' navigation history available to advertisers is proposed. For a given advertiser, we represent as a graph the partial navigation history of users who interacted with any element of a campaign such as clicking on text ads or banners, visiting a page of the advertiser's website containing a web tracker, etc. We modeled the problem assuming that an increase of traffic at one vertex implies a decrease for all vertices sharing the same parent(s) and vice versa. Based on that, we reorganize the flow within the graph to maximize the expected profits from conversions tracked online.

To solve the problem, we designed five algorithms. Four of them are based on the tabu search algorithm. In the most basic one, we change the position of one keyword by one unit then recalculate the flow using the functions mentioned earlier. The keyword which position is to be changed is chosen so that the change is the most profitable but respects the budget. Changing the position of that keyword then becomes tabu for a given number of iterations. The three other tabu algorithms are variations of the basic one. In the first variation, we allow the position of the chosen keyword to be changed by more than one unit. In the second, the change can lead to expenses greater than the budget. In the last one, both variations are combined.

The fifth algorithm is a greedy one. During every iteration, for each keyword, we determine whether keeping it at the same position or changing its position by one unit is more profitable. If changing is chosen, we then determine the number of units that makes the change the most

profitable. Exceeding the budget is allowed with a penalty. This algorithm turned out to be the fastest with results similar in value to those of the tabu algorithms.

The profit of a campaign is only one of the possible measures of its efficiency. In our second article, we analyze how different changes impact the campaign management by comparing the results obtained with different objective-functions. The first change analyzed was the use of a different objective-function. We compared the values of the performance indicators obtained when optimizing each one of them. The analysis revealed that the profit and the costs are the most sensitive to change. Also, visits maximization has the most negative impact than other objective-functions on performance indicators other than visits.

The second change we tested was the variation of budget. Profit maximization was not very sensitive to that. For the other objective-functions, it had more impact but $\pm 10\%$ or 20% did not necessarily translate to a variation of $\pm 10\%$ or 20% in value for any of them.

The last change was the variation of the cost-per-click. We randomly changed the cost-per-click of every keyword and compared the values of the performance indicators before and after the change. For profit maximization, less than 7% of the keywords saw their best positions change by one unit and the positions of other keywords did not change at all. The impact was more significant for the other objective-functions.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xiii
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
LISTE DES ANNEXES	xv
CHAPITRE 1 INTRODUCTION	1
1.1 Concepts de base	1
1.1.1 Types de publicité	1
1.1.2 Fonctionnement des campagnes de publicité en ligne	2
1.1.3 Informations disponibles	5
1.2 Éléments de la problématique	5
1.2.1 Modèles d’attribution des conversions	6
1.2.2 Stratégies d’optimisation de campagne	7
1.3 Objectifs de recherche	7
1.4 Plan de la thèse	8
CHAPITRE 2 REVUE DE LITTÉRATURE	9
2.1 Performance d’une campagne de marketing interactif	9
2.1.1 Optimisation d’enchère	9
2.1.2 Optimisation du budget	10
2.2 Prédiction en marketing interactif	11
2.2.1 Lien entre positions des annonces, taux de clic et coût par clic	11
2.2.2 Prédiction de clics	11

2.2.3	Analyse du comportement des utilisateurs	12
CHAPITRE 3	MODÈLE	13
3.1	Graphe	13
3.2	Modélisation du problème	15
3.2.1	Notations	15
3.2.2	Construction du graphe	17
3.2.3	Conservation de flot	24
3.2.4	Contraintes de budget	25
3.2.5	Modèle proposé	26
3.3	Un exemple	27
3.4	Autres fonctions-objectifs	30
CHAPITRE 4	ALGORITHMES	32
4.1	Assignation des positions	33
4.1.1	Recherche taboue	33
4.1.2	Algorithme glouton	41
4.2	Résolution de système d'équations linéaires : Méthodes itératives	42
4.2.1	Méthode de Jacobi	42
4.2.2	Méthode de Gauss-Seidel	43
4.2.3	Méthode de Gauss-Seidel optimisée	44
CHAPITRE 5	TESTS EXPÉRIMENTAUX	46
5.1	Première série de tests	46
5.2	Deuxième série de tests	50
5.2.1	Comparaison des résultats avec différentes fonctions-objectifs	51
5.2.2	Comparaison des résultats avec différents budgets	60
5.2.3	Comparaison des résultats avec différents <i>CPCs</i>	64
CHAPITRE 6	CONCLUSION	70
6.1	Synthèse des travaux	70
6.2	Limitations de la solution proposée	71
6.3	Améliorations futures	71
RÉFÉRENCES	73
ANNEXES	76

LISTE DES TABLEAUX

Tableau 5.1	Caractéristiques des instances	47
Tableau 5.2	Temps de calcul de l'algorithme glouton avec quatre différentes méthodes de résolution de systèmes d'équations linéaires	48
Tableau 5.3	Comparaisons du glouton aux quatre tabous combinés à la méthode Gauss-Seidel optimisée	50
Tableau 5.4	Caractéristiques des instances	50
Tableau 5.5	Changement d'objectif : Comparaison entre Tabou ₂ et Glouton	52
Tableau 5.6	Valeur des différentes mesures de performance à la meilleure solution trouvée pour la maximisation de profit	53
Tableau 5.7	Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de revenu comparativement à la maximisation de profit	55
Tableau 5.8	Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de conversion comparativement à la maximisation de profit	56
Tableau 5.9	Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de visite comparativement à la maximisation de profit	56
Tableau 5.10	Autres variations de valeur des mesures de performance à la meilleure solution trouvée	57
Tableau 5.11	Différence entre la meilleure valeur de chaque mesure obtenue avec différents scénarios et celle obtenue avec le scénario initial	59
Tableau 5.12	Différence entre la meilleure valeur de chaque mesure obtenue avec différents scénarios et celle obtenue avec le scénario initial en utilisant l'ordre lexicographique	60
Tableau 5.13	Valeur à la meilleure solution trouvée pour chaque fonction-objectif et pourcentage du budget dépensé à chaque solution	61
Tableau 5.14	Comparaison du profit dans le cas d'une maximisation de profit avec différents budgets	61
Tableau 5.15	Comparaison du revenu dans le cas d'une maximisation de revenu avec différents budgets	62
Tableau 5.16	Comparaison du nombre de conversions dans le cas d'une maximisation de conversion avec différents budgets	63

Tableau 5.17	Comparaison du nombre de visites dans le cas d'une maximisation de visite avec différents budgets	64
Tableau 5.18	Sommaire des variations des positions des mots-clés	66
Tableau 5.19	Variation de la proportion de mots-clés à toutes les positions à la meilleure solution trouvée après la variation des <i>avgCPCs</i>	68
Tableau 5.20	Variation de l' <i>avgCPC</i> pour chaque variation de position	69
Tableau B.1	Proportion de mots-clés à chaque position à la meilleure solution trouvée avant et après la variation des <i>avgCPCs</i> - Graphe 1 à 5	79
Tableau B.2	Proportion de mots-clés à chaque position à la meilleure solution trouvée avant et après la variation des <i>avgCPCs</i> - Graphe 6 à 10	80
Tableau B.3	Variation de l' <i>avgCPC</i> pour chaque variation positive de position . .	81
Tableau B.4	Variation de l' <i>avgCPC</i> pour chaque variation négative de position . .	82

LISTE DES FIGURES

Figure 3.1	Exemple construction de graphe : Chemin 1	22
Figure 3.2	Exemple construction de graphe : Chemin 2	22
Figure 3.3	Illustration de l'assignation de la position \bar{p} à un mot-clé	29
Figure 5.1	Amplitude de variation de l' <i>avgCPC</i> selon la position initiale d'un mot-clé : cas d'un mot-clé dont l' <i>avgCPC</i> initial est de 2\$ à la position 10	65

LISTE DES SIGLES ET ABRÉVIATIONS

CPC	Coût par clic : montant facturé à un annonceur pour chaque clic sur une annonce donnée
CTR	"Click-Through-Rate" : proportion d'utilisateurs qui cliquent sur une annonce donnée lorsqu'elle est affichée à leur écran
KPI	Indicateur de performance : mesure servant à évaluer la performance d'une campagne en général ou d'une annonce en particulier

LISTE DES ANNEXES

ANNEXE A	GLOSSAIRE	76
ANNEXE B	TABLEAUX COMPLETS	79

CHAPITRE 1 INTRODUCTION

Dans le cadre d'une campagne de publicité en ligne, une personne ou une entreprise dite annonceur crée une annonce publicitaire qu'il cherche à afficher sur un site Web tiers dit afficheur. Il existe plusieurs types d'annonces et pour chaque type, il y a un ou plusieurs réseaux de publicité adaptés. Généralement, plusieurs annonceurs convoitent le même espace d'affichage pour leurs annonces. Pour choisir celle qui sera affichée, l'afficheur met l'espace aux enchères. Chaque annonceur dit à l'afficheur le montant maximal qu'il est prêt à payer pour l'espace (*maxCPC*) sans connaître l'enchère des autres. Généralement, celui avec l'enchère la plus élevée l'emporte et il sera facturé soit à chaque millier d'affichages (encore appelées impressions) de son annonce, soit à chaque fois qu'un utilisateur cliquera sur celle-ci. Une telle enchère est tenue à chaque fois qu'un utilisateur ouvre le site d'un afficheur.

Chaque annonceur a un certain nombre de contraintes à respecter telles qu'un budget quotidien, hebdomadaire ou mensuelle, une durée minimale de campagne, un nombre minimum d'impressions, etc. L'objectif à atteindre est appelé une conversion. La notion de conversion varie selon l'annonceur. Il peut s'agir d'achat par les utilisateurs, de leur inscription à un bulletin d'informations ou d'une simple visite sur le site de l'annonceur.

Il y a encore quelques années, il n'était pas possible d'avoir accès à l'historique de navigation des utilisateurs. Il était alors impossible de déterminer ce qui influençait le cheminement de l'utilisateur et le menait à la conversion. Ces données étant désormais accessibles, nous pouvons aujourd'hui concevoir un modèle d'optimisation de campagne pour aider les annonceurs à améliorer leurs stratégies.

1.1 Concepts de base

1.1.1 Types de publicité

On distingue plusieurs types de publicité sur internet. On compte principalement le type :

- textuel : affiché par les moteurs de recherche, il consiste en un paragraphe de quatre lignes environ avec un titre, une description d'une ou deux lignes et une adresse URL. Chez Google et Bing, on les distingue des résultats dits organiques par le mot "Ad" affiché à côté de l'adresse URL ;
- graphique : affiché sur un réseau d'affichage, bien qu'il puisse arriver que le contenu soit textuel, il s'agit le plus souvent d'images statiques ou de courtes animations ;
- vidéo : affiché parfois sur les réseaux d'affichage mais surtout sur les réseaux sociaux

de partage de vidéos tels que YouTube. La vidéo de la publicité est diffusée avant celle que l'utilisateur a sélectionnée et peut durer de quelques dizaines de secondes à plusieurs minutes. Les vidéos peuvent être les mêmes que celles diffusées sur les chaînes de télévision ou avoir un contenu adapté à la navigation sur internet ;

- *rich media* : affiché généralement sur les réseaux d'affichages, il s'agit de graphiques ayant des fonctionnalités particulières tels que déclencher un appel, ajouter un événement au calendrier de l'utilisateur, etc.

Cliquer sur une publicité, peu importe son type, conduit généralement l'utilisateur sur un site Web choisi par l'annonceur. Néanmoins, dans le cas des *rich media*, l'action de la publicité peut varier selon le but visé par l'annonceur.

Dans le cadre de cette thèse, nous nous sommes limités aux publicités textuelles. Toutefois, une campagne publicitaire en ligne peut regrouper plus d'un type. Il est donc important de comprendre le fonctionnement de chaque réseau ainsi que la dynamique de l'ensemble.

1.1.2 Fonctionnement des campagnes de publicité en ligne

Moteurs de recherche

Sur les moteurs de recherche, chaque annonceur crée ses annonces textuelles ; il peut en créer plusieurs par campagne et les regrouper selon ce qu'elles ont en commun. Par exemple, dans une campagne de promotion d'une marque, chaque groupe d'annonces (*ad group*) peut servir à promouvoir un type de produits, une valeur de l'entreprise, etc. Les annonces appartenant au même groupe sont souvent des variantes d'une même annonce. On appelle alors chaque variante un créatif.

Pour chaque groupe d'annonces, l'annonceur fournit au moteur de recherche une liste de mots-clés. Ces mots-clés sont des mots et expressions auxquels seront comparés le texte que les utilisateurs ont tapé dans le moteur de recherche pour déterminer la pertinence de l'annonce. Pour chaque mot-clé, il est possible pour l'annonceur de préciser comment sera faite cette association. En théorie, chaque mot-clé d'un groupe d'annonces peut provoquer l'impression de n'importe quelle annonce du groupe ; et l'impression de toute annonce d'un groupe peut par conséquent être provoquée par n'importe quel mot-clé de ce groupe. En pratique, un moteur de recherche va progressivement déterminer quel créatif a le plus de succès avec un mot-clé et, éventuellement, un mot-clé ne provoquera l'apparition que d'un créatif alors qu'un créatif pourra être lié à plusieurs mots-clés.

Le texte que les utilisateurs tapent dans le champ prévu à cet effet dans le moteur de recherche est ce qui est appelé une requête. Pour chaque utilisateur, la requête est comparée par le moteur de recherche à tous les mots-clés des annonceurs pour faire une première sélection des

annonceurs. Pour la suite, considérons que le mot-clé est "*appartement à louer à Montréal*"; les options de comparaison sont les suivantes :

- mot-clé exact (*exact match*) : l'utilisateur doit entrer exactement le mot-clé. Par exemple, si la requête est "*appartement à louer*" ou "*grand appartement à louer à Montréal*", l'annonceur ne sera pas sélectionné;
- expression exacte (*phrase match*) : l'utilisateur doit entrer exactement le mot-clé mais peut ajouter d'autres mots avant ou après. Ainsi, si la requête est "*grand appartement à louer à Montréal*", l'annonceur sera sélectionné mais si la requête est juste "*appartement à louer*", l'annonceur ne sera pas sélectionné;
- requête large (*broad match*) : si le mot-clé compte plus d'un mot, l'utilisateur doit se servir d'au moins un de ces mots. Si la requête est "*maison à louer*" ou "*appartement à vendre*", l'annonceur sera donc sélectionné;
- expression négative (*negative match*) : aucun des mots commençant par un tiret dans le mot-clé ne doit figurer dans la requête. Considérons ici que le mot-clé est "*appartement à louer à Montréal -vendre*" (le mot "*vendre*" est donc un mot que l'annonceur rejette). Dans ce cas, si la requête est "*appartement à vendre*", l'annonceur ne sera pas sélectionné. Les expressions négatives ne sont jamais utilisées seules, elles sont toujours associées à d'autres mots-clés; on ne peut donc pas avoir uniquement "*-appartement*" comme mot-clé;
- requête large modifiée (*broad match modifier*) : il s'agit d'une requête large avec des impératifs. L'annonceur précise le ou les mots qui doivent être présents dans la requête. Par exemple, si les mots "*appartement*" et "*Montréal*" sont obligatoires, la requête "*appartement à louer*" ne suffira pas à sélectionner l'annonceur mais la requête "*appartement à vendre à Montréal*" permettra de le sélectionner.

Parmi les annonceurs sélectionnés, il faut maintenant les classer pour déterminer lesquels seront affichés et dans quel ordre. Pour faire ce classement, la majorité des moteurs de recherche se servent de l'algorithme *GSP (Generalized Second-Price)* qui consiste à classer les annonceurs en ordre décroissant de valeur d'enchère et ceux-ci sont facturés non pas le montant de leur enchère mais plutôt le montant minimum qu'il faut pour dépasser l'enchère du plus proche compétiteur. "Par exemple, si l'annonceur #1 mise 2,00\$ par clic et l'annonceur #2 mise 1,00\$ par clic, l'annonceur #1 paiera seulement 1,01\$ pour chaque clic qu'il recevra" Quinn (2011). Il est important de noter que bien que l'annonceur soit facturé pour chaque clic sur ses annonces, l'enchère elle, est sur le mot-clé.

Pour améliorer l'expérience de l'utilisateur, les moteurs de recherche tiennent parfois compte d'autres critères. Par exemple, Google utilise l'indice de qualité des annonces respectives des

annonceurs. Cet indice est une mesure qui reflète la performance passée de l'annonce. La manière dont il est calculé est inconnue des annonceurs et propre à chaque moteur de recherche. Aussi, ces derniers sont classés en fonction du produit de cet indice et de l'enchère : plus il est élevé, meilleur sera le classement de l'annonceur. La position la plus haute sur la première page de résultats a pour indice 1 et plus l'indice est grand, plus l'annonce est basse dans la page. Nous ne considérons ici que les dix premières positions qui sont toutes sur la première page. L'impact des positions subséquentes est considéré négligeable.

Les annonces publicitaires ainsi choisies et classées apparaissent sur la page de résultats avec ce qu'on appelle des annonces ou résultats organiques. Ce sont des liens affichés gratuitement sur un moteur de recherche qui ont un format similaire à celui des annonces publicitaires mais qui sont le fruit de ce qu'on appelle le référencement naturel. Il s'agit d'une méthode d'indexation du contenu d'un site web. Les moteurs de recherche le font en utilisant des mots-clés ou des métadonnées ce qui leur permet d'identifier des contenus pertinents selon la requête d'un utilisateur. L'optimisation de l'exploitation du référencement naturel plus généralement appelé optimisation du moteur de recherche (*search engine optimization* ou *SEO*) est un sujet de recherche majeur en marketing interactif que nous ne traiterons pas dans cette thèse.

Réseaux d'affichage

Les réseaux d'affichage consistent en un ensemble de sites Web qui offrent des espaces publicitaires sur leurs sites, en échange d'un revenu Quinn (2011). Les annonceurs peuvent choisir les sites sur lesquels ils veulent que leurs publicités soient affichées mais également exclure ceux sur lesquels ils ne veulent pas qu'elles soient affichées. La sélection des annonceurs se fait d'une façon similaire à celle sur les moteurs de recherche, de même que la facturation. Une différence notable est qu'il est également possible de facturer les annonceurs non pas par clic mais par impression. Ainsi, l'annonceur sera facturé chaque fois qu'un créatif de la campagne sera affiché à l'écran d'un utilisateur.

Sur les réseaux sociaux, les afficheurs possèdent beaucoup plus d'informations concernant les utilisateurs ce qui aide les annonceurs à atteindre le public-cible. D'ailleurs, les annonceurs peuvent donner de nombreuses précisions sur le public qu'ils cherchent à rejoindre. De plus, les utilisateurs peuvent partager des publicités avec leurs amis.

Sur les deux types de réseaux, pour éviter que les montants facturés soient trop élevés ou que le budget de la campagne ne s'épuise trop vite, l'annonceur peut fixer un budget périodique (quotidien, hebdomadaire ou mensuel). Une fois le budget atteint, la campagne est suspendue

jusqu’au début de la prochaine période.

1.1.3 Informations disponibles

Aujourd’hui, à moins que l’utilisateur ait interdit aux sites Web de le faire à travers son navigateur, il est possible d’obtenir certains détails de son historique de navigation. Ces détails incluent sans s’y limiter :

- le *cookie* qui permet d’identifier l’utilisateur ;
- l’adresse URL du site d’origine ;
- la date et l’heure à la seconde près à laquelle l’utilisateur a cliqué sur l’annonce ;
- le type d’annonce sur lequel l’utilisateur a cliqué ;
- le réseau sur lequel l’annonce a été affichée ;
- une information permettant de déterminer si l’utilisateur a cliqué sur une annonce publicitaire ou sur une annonce organique ;
- le rôle qu’a joué le site sur le cheminement de l’utilisateur. On distingue trois rôles :
 - initiateur (*introducer*) : premier événement qui a introduit l’utilisateur à la campagne. Il peut être entre autres un clic sur une annonce ou une impression ;
 - influenceur (*influencer*) : événement qui décrit le cheminement de l’utilisateur sur une période de temps pouvant couvrir plusieurs jours ;
 - fermeture (*closer*) : événement qui indique l’issue du parcours (une perte ou une conversion selon le cas).

L’ensemble de ces informations représente ce qu’on appelle le chemin de conversion de l’utilisateur et ce, que l’historique se termine ou non par une conversion. Plus de détails sur les chemins de conversion sont fournis dans la section 3.1 de cette thèse.

1.2 Éléments de la problématique

Lorsqu’un utilisateur tape une requête dans un moteur de recherche, pendant le chargement de la page des résultats, une enchère a lieu pour déterminer quelles annonces textuelles seront affichées et dans quel ordre. Le texte tapé par l’utilisateur doit pouvoir être associé à au moins un des mots-clés choisis par un annonceur pour qu’il soit éligible à participer à cette enchère. Il est reconnu (Brooks (2004a,b); Yao et Mela (2011)) que plus une annonce est affichée haut sur la page, mieux elle est classée, plus l’annonceur devra payer pour chaque clic sur ladite annonce et plus les utilisateurs cliqueront dessus. Toutefois, le montant facturé par

le moteur de recherche à un annonceur pour chaque clic ne dépasse jamais l'enchère de cet annonceur et le montant total de la facture ne doit pas dépasser son budget. Les annonceurs doivent donc établir une stratégie d'enchère qui leur permettra de maximiser la performance de leurs campagnes tout en respectant leur budget. Celle-ci se mesure de plusieurs façons selon l'objectif poursuivi par l'annonceur. Les mesures de performance comptent, sans s'y limiter :

- le nombre de conversions : cette mesure est particulièrement utile pour compter le nombre de ventes ou le nombre de nouvelles inscriptions à la liste de diffusion ;
- les revenus générés par la campagne : cette mesure permet de mesurer le volume des ventes effectuées grâce à la campagne ;
- le nombre de visites sur certaines pages du site de l'annonceur : celle-ci peut être utilisée comme mesure de visibilité ;
- les profits : c'est la différence entre les revenus générés par la campagne et les coûts de la campagne donc les profits supplémentaires obtenus grâce à la campagne.

Quel que soit la mesure choisie par l'annonceur, les questions qu'il est logique de se poser sont : "Quels sont les éléments de la campagne (parmi ceux présentés à la section 1.1.3) qui contribuent le plus à mon objectif?" et "Combien miser pour chacun d'eux?". Pour répondre à la première question, les annonceurs ont recours aux modèles d'attribution des conversions. Pour la seconde, malheureusement, ils n'ont accès qu'à une quantité limitée d'informations pour déterminer une stratégie d'optimisation de la campagne.

1.2.1 Modèles d'attribution des conversions

Il existe cinq modèles standards d'attribution de conversions dans l'industrie :

- dernière interaction (*last interaction*) : tout le crédit ou poids de la conversion est attribué à la dernière interaction la précédant. C'est le modèle de base ;
- première interaction (*first interaction*) : tout le crédit de la conversion est attribué à la première interaction du chemin ;
- linéaire (*linear*) : tout le crédit de la conversion est réparti de façon égale sur toutes les interactions du chemin de conversion de l'utilisateur ;
- décroissance dans le temps (*time decay*) : plus l'interaction de la campagne est récente, plus son poids est élevé ;
- basé sur la position (*position based*) : une majorité du crédit de la conversion est donnée à la première interaction et à la dernière de façon égale et le reste est réparti de façon égale sur les autres éléments du chemin de conversion de l'utilisateur.

Google AdWords offre également un modèle basé sur les données pour les annonceurs qui en ont accumulées suffisamment. Un défaut commun aux six modèles est qu'ils se limitent à un réseau, le moteur de recherche. De plus, les cinq modèles standards représentent rarement la réalité. En effet, les chemins des utilisateurs sont rarement identiques. Il est très probable qu'un événement occupe une position différente dans les différents chemins des utilisateurs. Il aurait donc différents poids selon le chemin observé. Dans cette thèse, nous ne présentons pas un modèle d'attribution mais plutôt un modèle d'optimisation qui tient compte de l'historique des utilisateurs et donc, implicitement, du modèle d'attribution le plus approprié pour l'annonceur observé.

1.2.2 Stratégies d'optimisation de campagne

Il existe de nombreux travaux de recherche afin d'améliorer la performance des campagnes publicitaires en ligne. Quelques-uns d'entre eux sont présentés dans la revue de littérature au chapitre 2. Toutefois, beaucoup d'annonceurs dépendent encore abondamment des rapports des réseaux pour déterminer sur quels mots-clés miser et le montant de la mise. Les modèles présentés dans la littérature servent généralement pour un seul objectif ou ne tiennent pas compte des historiques des utilisateurs, ni des interactions entre les éléments d'une campagne, ni entre différentes campagnes ; certains ne tiennent même pas compte des contraintes de budget.

Dans cette thèse, le modèle de programmation mathématique proposé peut être utilisé avec différents objectifs de campagne et tient compte des historiques des utilisateurs ainsi que des contraintes de budget. Les historiques permettent de représenter les interactions des utilisateurs avec les différents éléments de la campagne (voir section 1.1.3) et, s'ils ne se limitent pas à une campagne, les interactions entre les campagnes.

1.3 Objectifs de recherche

Notre projet a été réalisé en partenariat avec la compagnie Acquisio qui développe et fournit des logiciels d'optimisation de campagnes publicitaires par l'acquisition de nouvelle clientèle. La problématique présentée au début du projet est que chaque annonceur a accès depuis peu aux historiques en lien avec ses campagnes et souhaiterait s'en servir pour améliorer leur performance. L'objectif poursuivi par ce travail est donc de développer un modèle qui permet de maximiser la performance de toute campagne publicitaire avec les informations disponibles aux annonceurs. À travers cette thèse, nous cherchons à atteindre les objectifs suivants :

1. Modéliser l'ensemble des éléments des campagnes d'un annonceur, permettant ainsi de tenir compte des interactions entre ces éléments ;
2. Déterminer la meilleure position possible pour chaque combinaison mot-clé - annonce tout en respectant le budget et ce, pour tout objectif de campagne mesurable ;
3. Analyser l'impact d'un changement de budget, de coût-par-clic (*CPC*) ou d'objectif sur la performance d'une campagne.

La principale difficulté de ce projet réside dans la modélisation. En effet, le modèle doit tenir compte de toutes les informations disponibles, établir un lien cohérent entre elles et permettre d'évaluer l'impact de tout changement. Une fois ce modèle conçu, il faudra trouver un algorithme de résolution qui produit des résultats de bonne qualité dans un délai raisonnable.

1.4 Plan de la thèse

Le chapitre 2 présente une revue de la littérature. Dans le chapitre 3, dont l'essentiel du contenu est extrait de l'article par Azeuli et al. (2017) soumis pour publication dans un journal, un modèle de programmation mathématique est proposé. Au chapitre 4, différents algorithmes de résolution ainsi que divers tests d'impact sont présentés, testés et analysés. En effet, le modèle présenté au chapitre 3 est pour le cas de la maximisation du profit d'une campagne. Toutefois, le profit est loin d'être la seule mesure de performance d'une campagne. Aussi, au chapitre 4, l'impact du changement de la fonction-objectif sur la meilleure solution trouvée est analysé, et également celui du changement de budget ainsi et du changement de *CPC*. Une partie des résultats qui y sont présentés sont également tirés de l'article par Azeuli et al. (2017). Pour finir, vient la conclusion au chapitre 6 et les annexes. Un glossaire avec les définitions des termes-clés utilisés tout au long de cette thèse se trouve à l'annexe A.

CHAPITRE 2 REVUE DE LITTÉRATURE

Il existe de nombreux sujets de recherche en marketing interactif. Beaucoup s'attardent sur la dynamique d'enchères entre les annonceurs ou abordent les problèmes sous la perspective des afficheurs. Dans cette revue, sont regroupés, par thématique, des travaux qui touchent à l'amélioration de la performance d'une campagne et/ou à la prédiction de certaines mesures de performance.

2.1 Performance d'une campagne de marketing interactif

En marketing interactif, une problématique-clé est l'amélioration des performances des campagnes. Pour traiter de ce thème, les chercheurs ont abordé plusieurs approches. En voici quelques-unes.

2.1.1 Optimisation d'enchère

Le problème d'optimisation des enchères a pour but de déterminer combien miser pour avoir l'annonce affichée à la position la plus profitable. Borgs et al. (2007) ont développé une méthode pour trouver les enchères de chaque annonceur et chaque mot-clé pour maximiser leur utilité. Mais cela est fait pour plusieurs annonceurs convoitant les mêmes positions. Les annonceurs ne peuvent pas utiliser cette méthode, car ils n'ont pas accès aux données de la compétition. Even Dar et al. (2009) présente un travail se concentrant sur l'association de type "*broad match*". Toutefois, ils supposent implicitement que la position d'une annonce n'a pas d'influence sur le nombre de clics dessus, car elle n'est pas modélisée dans leur problème. Borgs et al. (2007) et Even Dar et al. (2009) utilisent la maximisation du profit pour optimiser les positions des annonces. Borgs et al. (2007) ont conçu une heuristique qui égalise le retour-sur-investissement de tous les mots-clés, alors que Even Dar et al. (2009) ont développé un algorithme basé sur le problème de coupe minimale en théorie des graphes pour identifier un sous-ensemble de requêtes qui remportent l'enchère avec une utilité maximale. Desai et al. (2014) analysent l'impact pour un annonceur de miser sur sa propre marque ainsi que sur celle d'un compétiteur. Le but visé étant de présenter son produit aux utilisateurs intéressés par celui de la compétition. Skiera et Abou Nabout (2013) ont créé un modèle pour déterminer l'enchère optimale sur un mot-clé qui permet de maximiser le profit espéré. Le modèle est conçu pour être utilisé dans un contexte d'enchère en temps réel et ne prend pas en compte les contraintes de budget ni l'impact de cette décision sur le reste de la campagne. D'ailleurs, la

seule contrainte prise en considération est que l'enchère ne doit pas dépasser le montant que cela coûterait de placer l'annonce en première position. Ils ont utilisé les formules proposées par Feng et al. (2007) pour estimer le taux de clic de l'annonce et le *CPC* du mot-clé. Zhou et al. (2008) ont formulé le problème d'optimisation d'enchères comme un problème de sac à dos où l'annonceur doit choisir un emplacement à cibler à un moment donné dans le but de maximiser soit ses revenus ou ses profits tout en ignorant les mises des autres annonceurs. Ils ne considèrent pas qu'il puisse y avoir un lien entre les enchères alors que, dans cette thèse, les historiques des utilisateurs sont pris en compte pour simuler l'impact potentiel de chaque décision sur le reste de la campagne. Zhang et al. (2012) ont proposé un modèle pour résoudre les problèmes d'optimisation d'enchères et d'allocation de budget pour maximiser le revenu des annonceurs.

2.1.2 Optimisation du budget

Dans le problème d'optimisation des budgets, les annonceurs veulent maximiser le nombre de clics sur les annonces tout en respectant leur budget. Archak et al. (2010) abordent le problème d'optimisation de budget en utilisant le processus de décision markovien. Ils supposent que la probabilité qu'un utilisateur passe d'un état à un autre est fixée et constante alors que dans cette thèse, il est supposé que la probabilité dépende de la position de l'annonce. DasGupta et Muthukrishnan (2013) résolvent le problème d'optimisation de budget avec des méthodes d'optimisation stochastique. Ils définissent un scénario avec un nombre de clics espéré sur les annonces en lien avec chaque mot-clé et évaluent la probabilité que ce scénario se réalise. Leur objectif est de trouver l'ensemble de positions des annonces qui maximise la somme des nombres de clics espérés de tous les scénarios. L'approche utilisée dans cette thèse utilise une logique similaire mais avec un seul scénario. Feldman et al. (2007) optimisent également le budget. Dans leur modèle, ils souhaitent déterminer la meilleure valeur qu'il faut attribuer à la mise et non identifier la meilleure position d'une annonce comme dans cette thèse. Les variables dans leur modèle sont les mises alors qu'ici ce sont des positions d'annonces. Karande et al. (2013) proposent un modèle qui maximise la fonction-objectif choisie par l'annonceur dans le but d'améliorer son allocation de budget. Ils présentent également différents algorithmes qu'ils ont conçus pour déterminer si un annonceur devrait participer à une enchère à laquelle il est éligible. Prenant en compte des paramètres tels que le budget de chaque annonceur et la fonction-objectif choisie, les enchères passées et les taux de clic, ils choisissent quels annonceurs participeront à chaque enchère. Lu et al. (2015) analysent l'impact de la variation de budget sur les profits de l'annonceur mais aussi sur les revenus de l'afficheur. Yang et al. (2012) conçoivent une structure pour l'optimisation du budget pour en allouer une portion aux différents moteurs de recherche en minimisant le nombre de clics

perdus. Le travail de Zhang et al. (2012) rentre également dans cette catégorie.

2.2 Prédiction en marketing interactif

Une autre problématique majeure étroitement liée à la performance est celle de la prédiction de mesures de performance.

2.2.1 Lien entre positions des annonces, taux de clic et coût par clic

Les travaux qui traitent du lien entre positions des annonces, taux de clic et coût par clic cherchent à établir des formules permettant de déterminer la valeur de l'un en fonction d'un ou des deux autres. Brooks (2004a) indiquent que les annonces les plus haut placées obtiennent plus de clics et concluent que les positions les plus élevées coûtent plus cher (Brooks, 2004b). Yao et Mela (2011) sont également d'avis que les utilisateurs cliquent plus sur les annonces placées plus haut parmi les résultats de recherche. Quinn (2011) développe ce qu'il appelle des fonctions génériques. L'une d'entre elles sert à prédire le nombre de clics sur une annonce en fonction de la position à laquelle elle est affichée. L'autre fonction générique développée par Quinn utilise également les positions des annonces pour prédire leurs *CPC*. Une limite de ces fonctions est qu'elles ne tiennent pas compte d'autres paramètres ou variables pouvant influencer leurs résultats. Feng et al. (2007) ainsi que Kitts et Leblanc (2004) trouvent des formules similaires.

2.2.2 Prédiction de clics

La prédiction de clics, comme son nom l'indique, a pour but de prédire le nombre de clics qu'une annonce va recevoir. McMahan et al. (2013) traitent de la performance d'un algorithme d'apprentissage servant à estimer la probabilité qu'un utilisateur clique sur une annonce connaissant la requête à laquelle l'annonce a été associée et de sa mise à jour en temps réel. L'algorithme en question incorpore un gradient modifié où le pas représente le taux (ou la vitesse) d'apprentissage. Ils y mettent l'emphasis sur la performance informatique de l'algorithme plutôt que sur la qualité du résultat. Graepel et al. (2010) proposent également un algorithme d'apprentissage pour prédire le taux de clics. Cet algorithme implique l'utilisation de plusieurs attributs qui ne sont généralement pas accessibles aux annonceurs. Rutz et al. (2012) ont proposé un modèle pour prédire le taux de clics en utilisant les positions des annonces et les caractéristiques des mots-clés comme variables. Toutefois, les auteurs n'ont pas d'information sur l'historique des utilisateurs. De plus, ils affirment que certains paramètres ne sont pas significatifs sans se justifier.

2.2.3 Analyse du comportement des utilisateurs

En marketing interactifs, l'accès aux historiques de navigation partiels des utilisateurs a amené plus d'un à chercher à comprendre ce qui motive les utilisateurs à interagir avec des annonces publicitaires. Dans cette thèse, la théorie des graphes est utilisée pour représenter les interactions des utilisateurs. Cette idée a été explorée plus d'une fois dans le contexte du marketing en ligne. Nasraoui et Krishnapuram (2002) proposent un algorithme de partitionnement permettant de lier les liens URL en définissant un contexte de navigation et d'extraire des historiques d'un utilisateur un profil de navigation. Cet algorithme permettra de définir des contextes de navigation des utilisateurs pour mieux les comprendre et mieux les regrouper mais peut nécessiter une quantité considérable de mémoire si les graphes sont trop grands. Plus tard, Nasraoui et al. (2003) s'inspirent du fonctionnement du système immunitaire pour regrouper les utilisateurs ayant des profils similaires. Les données utilisées sont les historiques de navigation des utilisateurs. Li et al. (2009) cherchent à trouver les k chemins partiels (avec k fixé) les plus récurrents. Pour résoudre ce problème, ils se servent d'un algorithme permettant d'extraire des données historiques les chemins les plus utilisées. Une limite majeure est qu'il faille fixer le nombre de chemins qu'on veut obtenir. De plus, cet algorithme fait disparaître les chemins partiels qui commencent et se terminent par le même nœud.

CHAPITRE 3 MODÈLE

L'essentiel du contenu de ce chapitre est extrait de l'article par Azeuli et al. (2017) soumis pour publication dans un journal. Cet article est à la fin de cette thèse. Dans ce chapitre, sera présenté comment les données historiques fournies par notre partenaire industriel Acquisio ont été converties en graphe, comment le graphe artificiel est généré. Ensuite, les contraintes du problème seront présentées et modélisées puis le reste du modèle mathématique sera présenté et expliqué. Ces données présentées à la section 1.1.3 permettent de représenter les interactions des utilisateurs avec les annonces présentes dans leurs chemins de conversion.

3.1 Graphe

Les données qui nous ont été fournies par notre partenaire industriel sont les chemins de conversion, le nombre de clics par jour sur chaque annonce, les *CPCs* moyens (*avgCPC*) par jour des mots-clés et des bannières ainsi que les positions moyennes par jour des annonces. Dans les chemins de conversion, sont représentés les historiques de navigation partiels des utilisateurs. Plus précisément, un chemin de conversion contient chaque interaction qu'un utilisateur a eue avec les éléments d'une campagne. Ces éléments peuvent être une annonce, une bannière, l'adresse URL d'une page du site de l'annonceur ou une requête. Ainsi, chaque fois qu'un utilisateur utilise un moteur de recherche, il tape un texte dans le champ prévu à cet effet. Ce texte est la requête. Celle-ci va être associée à un mot-clé de l'annonceur et, s'il réussit à se classer dans l'une des dix premières positions et que son budget le permet, son annonce paraîtra sur la page. Et, si l'utilisateur clique sur l'annonce, dans le chemin de conversion, nous aurons la requête, le mot-clé auquel elle a été associée et l'annonce qui a été affichée avec la date et l'heure auxquelles l'utilisateur a cliqué ainsi que l'URL de la page de destination.

La requête de l'utilisateur va également provoquer l'apparition de liens qui ne sont pas des annonces publicitaires. C'est ce qu'on appelle des résultats organiques. Il est possible qu'au moins un de ces liens mène à une des pages du site de l'annonceur. Si l'utilisateur clique sur un lien qui mène à une des pages du site de l'annonceur, dans le chemin de conversion, nous aurons la requête et l'URL de la page en question avec la date et l'heure auxquelles l'utilisateur a cliqué. Il est également possible que la page soit dans un chemin de conversion sans que l'utilisateur ne soit passé par un moteur de recherche pour y parvenir. Pour cela, elle doit être munie d'un code servant à surveiller le trafic sur cette page.

Il est également possible que l'utilisateur soit tout simplement exposé à de la publicité sur la

page d’une tierce parti. Si cette publicité est une bannière de l’annonceur, que l’utilisateur clique dessus ou non, elle paraîtra dans son chemin de conversion avec la date et l’heure à laquelle cette bannière est apparue à l’écran de l’utilisateur.

Chaque utilisateur peut générer plus d’un chemin de conversion. Pour représenter ces chemins, nous avons créé un graphe dans lequel chaque élément est représenté par un nœud à l’exception des mots-clés et des annonces qui sont regroupés dans le même nœud. Dans cette thèse, un mot-clé peut être lié à une seule annonce textuelle mais une annonce peut être liée à plus d’un mot-clé. C’est pourquoi dans cette thèse, nous ferons référence à la paire mot-clé - annonce plutôt que de les désigner séparément. Nous regroupons donc un mot-clé et l’annonce dont il a provoqué l’apparition dans un même nœud que nous désignerons par le mot-clé. De plus, dans le graphe, il y a un arc (u, v) si et seulement si l’élément représenté par u est immédiatement suivi de l’élément représenté par v dans au moins un de ses chemins de conversion et le flot sur un arc est le nombre de fois que tout utilisateur est passé de l’événement représenté par u à l’événement représenté par v .

Au graphe ainsi construit, sont ajoutés quatre nœuds : la source s , le puits t , le nœud c seul successeur de tous les nœuds-conversions et le nœud-perte l qui est l’un des successeurs (si ce n’est le seul) de tous les nœuds, autres que des conversions, par lesquels se sont terminés des chemins de conversion. Ainsi, le flot sur un arc (s, v) est le nombre de chemins qui ont commencé par l’événement représenté par v , le flot sur (u, c) est le nombre de chemins qui se sont terminés par une conversion u , le flot sur (c, t) est le nombre de chemins convertis, le flot sur (u, l) est le nombre de chemins non convertis qui se sont terminés par l’événement représenté par u et le flot sur (l, t) est le nombre de chemins non convertis. De plus, par définition, c et l sont les seuls prédécesseurs de t . Un exemple est donné à la figure 3.3 de la section 3.3.

Le fait de représenter tous les chemins qu’ont parcourus des utilisateurs dans le même graphe permet d’obtenir des chemins qui, même s’ils ne paraissent dans aucun historique, sont possibles au moins en théorie. Cela permet d’évaluer les répercussions d’un changement apporté à un élément de la campagne sur les autres éléments représentés dans le graphe. En effet, sachant que le flot qui traverse le graphe ainsi construit correspond aux clics des utilisateurs, on comprend que toute variation de flot pourra être propagée dans le reste de ce graphe pour maintenir la conservation de flot.

Quinn (2011) et Skiera et Abou Nabout (2013) ont développé des formules liant la position d’une annonce au nombre de clics sur celle-ci ainsi qu’au *CPC* moyen du mot-clé qui a provoqué l’impression de cette annonce. Il en découle que tout changement de position d’une annonce aura un impact sur le *CPC* du mot-clé qui y est lié mais aussi sur le nombre de clics sur cette annonce et donc sur le flot traversant le nœud représentant le mot-clé et l’annonce.

La question qui se pose est donc : "à quelle position chaque annonce doit-elle être affichée pour optimiser la performance de la campagne ?"

3.2 Modélisation du problème

Nous avons formulé quelques hypothèses sur lesquelles est basé notre modèle :

- A1 On ne peut pas créer de flot supplémentaire : le flot à la source demeure constant. En effet, il ne nous est pas possible de prédire d'où ce flot pourrait provenir. Aussi, par souci de simplification, nous supposons que le flot à la source ne varie jamais. Il s'en suit que l'impact sur le reste du graphe se traduira strictement en une réorganisation du flot existant. Bien que cette hypothèse puisse sembler restrictive, il est possible de pallier ce problème en mettant à jour régulièrement le graphe et le flot qui y circule quand de nouveaux historiques sont disponibles.
- A2 Si la position d'une annonce est améliorée, cela générera plus de clics sur celle-ci en en prélevant des éléments qui partagent le(s) même(s) prédécesseur(s). Ceci est une conséquence de l'hypothèse A1 : une meilleure position mène à plus de clics mais comme nous ne générons pas de flot supplémentaire, ces clics en plus ne peuvent qu'être prélevés des éléments représentés par les nœuds-frères (nœuds ayant au moins un prédécesseur en commun). Dans ce travail, nous considérons que les autres annonceurs sont représentés par le nœud-perte l qui sera toujours parmi les nœuds-frères de toute paire mot-clé - annonce..
- A3 Si la position d'une annonce est détériorée, elle perdra des clics en faveur des éléments qui partagent le(s) même(s) prédécesseur(s). Ceci est également une conséquence de l'hypothèse A1 : une position moins bonne mène à moins de clics. Toutefois, ces clics ne peuvent pas tout simplement disparaître. Ils sont donc transférés aux éléments représentés par les nœuds-frères.

3.2.1 Notations

Pour la suite, nous utiliserons les notations suivantes.

Ensembles

- \mathcal{A} ensemble d'arcs
- \mathcal{V} ensemble de nœuds
- \mathcal{B} ensemble de nœuds représentant des bannières
- \mathcal{C} ensemble de nœuds représentant des conversions

\mathcal{K}	ensemble de nœuds représentant des mots-clés
\mathcal{Q}	ensemble de nœuds représentant des requêtes
\mathcal{W}	ensemble de nœuds représentant des pages du site du client
Γ_v^+	ensemble de nœuds-successeurs immédiats du nœud v
Γ_v^-	ensemble de nœuds-prédécesseurs immédiats du nœud v
\mathcal{P}	ensemble $\{1, \dots, \bar{p}\}$ de positions qui peuvent être assignées à un mot-clé. La position 1 attire le plus grand nombre de clics. Un mot-clé à la position \bar{p} est considéré comme n'étant pas affiché car la variation des nombres de conversions et de clics est trop faible au-delà de cette position. Pour la suite, nous avons fixé $\bar{p} = 11$.

Paramètres

β_{click}	coefficient d'ajustement du nombre de clics quand un mot-clé est affiché une position au-dessus
β_{cpc}	coefficient d'ajustement du CPC moyen quand un mot-clé est affiché une position au-dessus
B	budget de la campagne
B_{da}	budget pour la campagne sur le réseau d'affichage ($B_{da} < B$)
cpc_b	CPC moyen de la bannière $b \in \mathcal{B}$
cpc_k	CPC moyen initial du mot-clé $k \in \mathcal{K}$
δ	proportion de toutes les variations de flot attribuée à la perte
r_v	valeur d'une conversion $v \in \mathcal{C}$

Variables

$y_{k,p}$	$= \begin{cases} 1 & \text{si le mot-clé } k \text{ est placé à la position } p \\ 0 & \text{sinon} \end{cases}$
x_v	nouveau flot traversant le nœud $v \in \mathcal{V}$ quand de nouvelles positions sont assignées aux mots-clés
$x_{(v,w)}$	nouveau flot sur l'arc $(v, w) \in \mathcal{A}$

Notations supplémentaires

c	nœud représentant l'unique successeur de toutes les conversions
-----	---

s, t	nœuds représentant la source et le puits du graphe, respectivement
l	nœud représentant les pertes
p_k^{init}	position initiale du mot-clé $k \in \mathcal{K}$
n_v	flot initial traversant le nœud $v \in \mathcal{V}$
$n_{(v,w)}$	flot initial circulant sur l'arc $(v, w) \in \mathcal{A}$
$c_{k,p}$	CPC moyen estimé du mot-clé $k \in \mathcal{K}$ quand il est placé à la position $p \in \mathcal{P}$
$\Delta_{(v,k)}^p$	variation de flot estimée circulant sur l'arc (v, k) quand le mot-clé $k \in \mathcal{K}$ est placé à la position $p \in \mathcal{P}$
d_v	variation de flot estimée circulant sur l'arc liant le nœud v au nœud-perte l , quand de nouvelles positions sont assignées aux mots-clés
μ_v	nouveau flot estimé traversant le nœud $v \in \mathcal{V}$
$\mu_{(v,w)}$	nouveau flot estimé circulant sur l'arc $(v, w) \in \mathcal{A}$
$\alpha_{(v,w)}$	nouvelle proportion de flot allant de v à un successeur $w \in \Gamma_v^+$ of v .

3.2.2 Construction du graphe

Processus à partir des données réelles

Le partenaire industriel a fourni deux fichiers Excel. Le premier, noté \mathcal{E}_1 , contient les historiques des interactions des utilisateurs avec les campagnes d'un client donné; chaque ligne du fichier contient les informations d'un et un seul événement (voir section 1.1.3). Tous les événements sont triés en ordre chronologique du plus récent au plus ancien.

Le second fichier, noté \mathcal{E}_2 , contient diverses informations statistiques et pécuniaires concernant les campagnes du même client et couvre la même période de temps que le premier fichier. Les informations suivantes proviennent de ce fichier :

- la position moyenne;
- les coûts pour chaque élément;
- le coût par clic (CPC) moyen.

En une passe de chacun de ces fichiers, l'information nécessaire pour créer le graphe $\mathcal{G}(\mathcal{V}, \mathcal{A})$ représentant le problème est recueillie. Chaque nœud de \mathcal{V} représente un événement et chaque événement ne peut être représenté que par un seul nœud. Pour tout nœud v de \mathcal{V} , la liste de ses prédécesseurs Γ_v^- ainsi que celle de ses successeurs Γ_v^+ sont établies. De plus, le total des coûts générés par l'événement représenté par v , noté cpc_v , est extrait.

Un arc (i, j) de \mathcal{A} lie le nœud i de \mathcal{V} au nœud j de \mathcal{V} si et seulement si l'élément représenté par le nœud j suit immédiatement celui représenté par le nœud i dans l'historique d'au moins un utilisateur. n_i représente le nombre de fois que tout utilisateur a été exposé à l'événement

i et $n_{(i,j)}$ le nombre de fois que tout utilisateur est passé de l'événement i à l'événement j dans un de ses chemins.

Lors de la lecture du fichier \mathcal{E}_1 , s'il y a un délai entre deux interactions successives supérieur au seuil choisi ou si l'identifiant de la session change, on marque la fin du chemin et on en commence un nouveau.

Pseudocode Le pseudocode de construction du graphe se présente comme suit :

Algorithme 1 Construction du graphe

1: **Procédure** CONSTRUCTIONGRAPHE($\mathcal{E}_1, \mathcal{E}_2$) \triangleright Procédure pour construire le graphe à partir des fichiers des historiques et des données statistiques

Entrées : $\mathcal{E}_1, \mathcal{E}_2$

Sorties : $\mathcal{G}(\mathcal{V}, \mathcal{A})$

2: EXTRACTIONCHEMINS(\mathcal{E}_1)

3: EXTRACTIONCOUTS(\mathcal{E}_2)

4: **Fin Procédure**

1: **Procédure** EXTRACTIONCHEMINS(\mathcal{E}_1) \triangleright Procédure pour construire le graphe à partir du fichier des historiques

Entrées : \mathcal{E}_1

\triangleright Fichier Excel des historiques

Sorties : $\mathcal{G}(\mathcal{V}, \mathcal{A}), \mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{Q}, \mathcal{W}$

2: $\mathcal{A} \leftarrow \emptyset$

3: $\mathcal{V} \leftarrow \{s, c, l, t\}$

4: $\mathcal{B} \leftarrow \emptyset$

5: $\mathcal{C} \leftarrow \emptyset$

6: $\mathcal{K} \leftarrow \emptyset$

7: $\mathcal{Q} \leftarrow \emptyset$

8: $\mathcal{W} \leftarrow \emptyset$

9: $prevV \leftarrow s$

\triangleright Nœud précédent

10: $prevUser \leftarrow \text{nil}$

\triangleright Utilisateur précédent

11: $d \leftarrow$ délai maximum entre deux événements successifs d'un chemin

12: $path \leftarrow \text{nil}$

\triangleright Chemin courant

13: **Pour tout** événement e d' \mathcal{E}_1 **Faire**

14: $user \leftarrow \text{IP}(e)$ \triangleright Retourne l'adresse IP de l'appareil à partir duquel l'évènement a eu lieu

15: $v \leftarrow \text{VERTEX}(e)$ \triangleright Nœud représentant la valeur de l'élément pour l'évènement courant

```

16:      Si  $prevV = t$  alors
17:           $path \leftarrow nil$ 
18:           $prevV \leftarrow s$ 
19:          Incréments  $n_s$ 
20:      Fin Si
21:      Si  $\mathcal{V}$  contient  $v$  alors
22:          Incréments  $n_v$ 
23:      Sinon
24:          Ajouter  $v$  à  $\mathcal{V}$ 
25:          Si  $e$  est une annonce textuelle alors
26:              Fusionner  $prevV$  et  $v$ 
27:               $v \leftarrow prevV$ 
28:          Sinon
29:              Ajouter  $v$  au chemin  $p$ 
30:              Si  $e$  est une conversion alors
31:                  Ajouter  $v$  à  $\mathcal{C}$ 
32:              Sinon Si  $e$  est une bannière alors
33:                  Ajouter  $v$  à  $\mathcal{B}$ 
34:              Sinon Si  $e$  est un mot-clé alors
35:                  Ajouter  $v$  à  $\mathcal{K}$ 
36:              Sinon Si  $e$  est une page du site de l'annonceur alors
37:                  Ajouter  $v$  à  $\mathcal{W}$ 
38:              Fin Si
39:          Fin Si
40:      Fin Si
41:      Si  $prevV$  est différent de  $v$  ET  $prevUser = user$  ET le délai entre l'événement
    précédent et l'événement courant est plus petit que  $d$  alors
42:           $a \leftarrow ARC(prevV, v)$ 
43:          Si  $\mathcal{A}$  contient  $a$  alors
44:              Incréments  $n_a$ 
45:          Sinon
46:              Ajouter  $a$  à la liste  $\mathcal{A}$ 
47:          Fin Si
48:      Sinon Si  $prevUser$  est différent de  $user$  OU le délai entre l'événement précédent
    et l'événement courant est plus grand que  $d$  alors
49:          Si  $\mathcal{C}$  contient  $v$  alors
50:               $a \leftarrow ARC(v, c)$ 
51:              Incréments  $n_c$ 
52:          Si  $\mathcal{A}$  contient  $a$  alors
53:              Incréments  $n_a$ 
54:          Sinon
55:              Ajouter  $a$  à la liste  $\mathcal{A}$ 
56:          Fin Si

```

```

57:       $a \leftarrow \text{ARC}(c, t)$ 
58:      Incréments  $n_t$ 
59:      Si  $\mathcal{A}$  contient  $a$  alors
60:          Incréments  $n_a$ 
61:      Sinon
62:          Ajouter  $a$  à la liste  $\mathcal{A}$ 
63:      Fin Si
64:       $v \leftarrow t$ 
65:      Sinon Si  $prevUser$  est différent de  $user$  OU le délai entre l'événement précé-
        dent et l'événement courant est plus grand que  $d$  alors
66:           $a \leftarrow \text{ARC}(v, l)$ 
67:          Incréments  $n_l$ 
68:          Si  $\mathcal{A}$  contient  $a$  alors
69:              Incréments  $n_a$ 
70:          Sinon
71:              Ajouter  $a$  à la liste  $\mathcal{A}$ 
72:          Fin Si
73:           $a \leftarrow \text{ARC}(l, t)$ 
74:          Incréments  $n_t$ 
75:          Si  $\mathcal{A}$  contient  $a$  alors
76:              Incréments  $n_a$ 
77:          Sinon
78:              Ajouter  $a$  à la liste  $\mathcal{A}$ 
79:          Fin Si
80:           $v \leftarrow t$ 
81:      Fin Si
82:  Fin Si
83:   $prevV \leftarrow v$ 
84:   $prevUser \leftarrow user$ 
85:  Fin Pour
86:   $\mathcal{G} \leftarrow \mathcal{G}(\mathcal{V}, \mathcal{A})$ 
87: Fin Procédure

```

1: Fonction VERTEX(e)	▷ Crée un nœud et le retourne
Sorties : v	
2: $v \leftarrow e$	▷ Le nœud créé
3: $\Gamma_v^- \leftarrow \emptyset$	▷ Ensemble des prédécesseurs de v
4: $\Gamma_v^+ \leftarrow \emptyset$	▷ Ensemble des successeurs de v
5: $cpc_v \leftarrow 0$	▷ Coût-par-clic de v
6: $r_v \leftarrow 0$	▷ Revenu par visite de v
7: $n_v \leftarrow 1$	▷ Nombre de visiteurs de v
8: Retourner v	
9: Fin Fonction	

1: Fonction ARC(u, v)	▷ Crée un arc et le retourne
Sorties : a	
2: $a \leftarrow (u, v)$	▷ L'arc créé
3: $n_v \leftarrow 1$	▷ Nombre de visiteurs de a
4: Ajouter v à Γ_u^+	
5: Ajouter u à Γ_v^-	
6: Retourner a	
7: Fin Fonction	

1: Procédure EXTRACTIONCOUTS(\mathcal{E}_2)	▷ Met à jour les coûts pour tous les nœuds
Entrées : \mathcal{E}_2	▷ Fichier Excel des données statistiques
Sorties : cpc_v ($\forall v \in \mathcal{K} \cup \mathcal{B}$), r_v ($\forall v \in \mathcal{C}$)	
2: Pour tout événement e d' \mathcal{E}_2 Faire	
3: $v \leftarrow$ nœud de \mathcal{V} représentant l'événement e	
4: Si \mathcal{B} contient v alors	
5: $cpc_v \leftarrow$ coût de l'événement e	
6: Sinon Si \mathcal{K} contient v alors	
7: $cpc_v \leftarrow$ coût de l'événement e	
8: $p_v \leftarrow$ position moyenne initiale de l'événement e	
9: Sinon Si \mathcal{C} contient v alors	
10: $r_v \leftarrow$ revenu de l'événement e	
11: Fin Si	
12: Fin Pour	
13: Fin Procédure	

Exemple Considérons un historique avec les deux chemins suivants :

Chemin 1 : $q \rightarrow w_1 \rightarrow c_1$

Chemin 2 : $k \rightarrow w_1 \rightarrow b \rightarrow w_2 \rightarrow c_2$

Le graphe est construit comme suit :

Figure 3.1 Exemple construction de graphe : Chemin 1

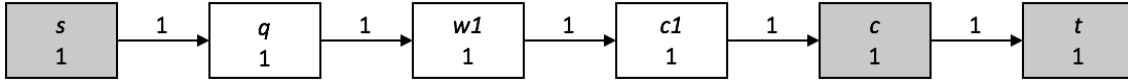
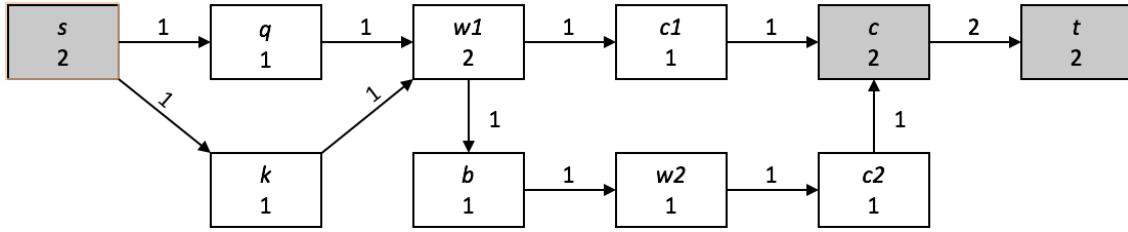


Figure 3.2 Exemple construction de graphe : Chemin 2



Processus de génération de graphe aléatoire

Pour générer un graphe aléatoire, un nombre de nœuds est donné ainsi qu'un nombre de chemins et une proportion pour quatre des cinq types de nœuds. Une longueur minimale (d'au moins 5 si on tient compte de s , c , l et t) et une longueur maximale des chemins sont fixées ainsi qu'un coût-par-clic minimal et un coût-par-clic maximal pour les mots-clés et pour les bannières. Par la suite le programme crée des nœuds de chaque type dans la quantité demandée en plus des nœuds s , c , l et t et génère des chemins en utilisant les nœuds générés, en priorité ceux qui n'ont pas encore été utilisés.

Pseudocode Le pseudocode de construction d'un graphe aléatoire se présente comme suit :

Algorithme 2 Construction du graphe artificiel

- 1: **Fonction** GRAPH($nbPaths, nbK, nbB, nbC, nbQ, nbW$) \triangleright Crée un graphe et le retourne
 - Sorties :** $\mathcal{G}(\mathcal{V}, \mathcal{A}), \mathcal{B}, \mathcal{C}, \mathcal{K}, \mathcal{Q}, \mathcal{W}$
 - 2: $minR \leftarrow$ revenu minimal d'une conversion
 - 3: $maxR \leftarrow$ revenu maximal d'une conversion
 - 4: $minC \leftarrow$ cpc minimal d'un mot-clé
 - 5: $maxC \leftarrow$ cpc maximal d'un mot-clé
 - 6: $minBC \leftarrow$ cpc minimal d'une bannière
 - 7: $maxBC \leftarrow$ cpc maximal d'une bannière
-

```

8:    $\mathcal{A} \leftarrow \emptyset$ 
9:    $\mathcal{B} \leftarrow nbB$  nœuds avec un flot nul et un coût par visite aléatoire compris entre  $minBC$ 
   et  $maxBC$ 
10:   $\mathcal{C} \leftarrow nbC$  nœuds avec un flot nul et un revenu par visite aléatoire compris entre  $minR$ 
   et  $maxR$ 
11:   $\mathcal{K} \leftarrow nbK$  nœuds avec un flot nul et un coût par visite aléatoire compris entre  $minC$ 
   et  $maxC$  et une position initiale aléatoire comprise entre 1 et  $\bar{p} - 1$ 
12:   $\mathcal{Q} \leftarrow nbQ$  nœuds avec un flot nul et
13:   $\mathcal{W} \leftarrow nbW$  nœuds avec un flot nul et
14:   $\mathcal{V} \leftarrow \{s, c, l, t\} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{K} \cup \mathcal{Q} \cup \mathcal{W}$ 
15:   $minPathLength \leftarrow$  longueur minimale d'un chemin
16:   $maxPathLength \leftarrow$  longueur maximale d'un chemin
17:   $i \leftarrow 0$  ▷ Compteur de chemins
18:  Pour  $i = 0 \rightarrow nbPaths$  Faire
19:     $path \leftarrow \text{nil}$ 
20:     $length \leftarrow$  un entier aléatoire compris entre  $minPathLength$  et  $maxPathLength$ 
21:     $prevV \leftarrow s$ 
22:    Incréments  $n_s$ 
23:     $v \leftarrow$  un nœud de  $\mathcal{B} \cup \mathcal{Q} \cup \mathcal{W}$ 
24:     $a \leftarrow \text{ARC}(prevV, v)$ 
25:    Incréments  $n_v$ 
26:    Ajouter  $a$  à la liste  $\mathcal{A}$ 
27:    Ajouter  $prevV$  et  $v$  à  $path$ 
28:    Tant que longueur de  $path$  est plus petite que  $length$  OU  $v$  n'est pas une conver-
   sion Répéter
29:       $prevV \leftarrow v$ 
30:      Si  $v \in \mathcal{B} \cup \mathcal{K} \cup \mathcal{Q}$  alors
31:         $v \leftarrow$  un nœud de  $\mathcal{W}$  (un qui n'a pas encore été utilisé s'il en reste)
32:      Sinon Si  $v \in \mathcal{W}$  alors
33:         $v \leftarrow$  un nœud de  $\mathcal{B} \cup \mathcal{K} \cup \mathcal{Q} \cup \mathcal{C}$  (un qui n'a pas encore été utilisé s'il en
   reste)
34:      Fin Si
35:       $a \leftarrow \text{ARC}(prevV, v)$ 
36:      Si  $\mathcal{A}$  contient  $a$  alors
37:        Incréments  $n_a$ 
38:      Sinon
39:        Ajouter  $a$  à la liste  $\mathcal{A}$ 
40:      Fin Si
41:    Fin Tant que
42:  Fin Pour
43: Fin Fonction

```

3.2.3 Conservation de flot

Supposons qu'un mot-clé $k \in \mathcal{K}$ change de position de p_k^{init} à p_k . Le flot (nombre de clics) dans le graphe doit être modifié selon les hypothèses A1, A2 et A3.

Soit un arc $(v, w) \in \mathcal{A}$ et $\mu_{(v,w)}$ le flot estimé circulant sur (v, w) suite à ces changements :

- si $w \notin \mathcal{K} \cup \{l\}$: le flot estimé $\mu_{(v,w)}$ sur (v, w) est égal au flot initial $n_{(v,w)}$ car nous nous limitons aux changements de position des annonces textuelles ;
- si $w \in \mathcal{K}$: on fixe $\mu_{(v,w)} = 0$ si $p_w = \bar{p}$ (l'annonce n'étant pas affichée, aucun utilisateur ne peut cliquer dessus), et on utilise la formule d'estimation par Quinn (2011) si $p_w \neq \bar{p}$. Plus précisément, on fixe

$$\mu_{(v,w)} = n_{(v,w)} (\beta_{click})^{p_w - p_w^{init}} \quad \text{pour tout } w \in \mathcal{K}, p_w \neq \bar{p}.$$

Ces estimations soulèvent le problème de conservation de flot. En effet, si le mot-clé k change de position et qu'il n'y a pas de mot-clé parmi ses ancêtres dans le graphe, alors le flot entrant dans tout nœud $v \in \Gamma_k^-$ ne peut pas varier, et le flot sur tout arc (v, k) doit aussi changer. Pour être sûr que les contraintes de conservation de flot soient respectées, les estimations précédentes sont traitées comme des demandes : chaque arc $(v, k) \in \mathcal{A}$ où $k \in \mathcal{K}$ demande que son flot actuel (nombre de clics) $n_{(v,k)}$ soit augmenté ou réduit à $\mu_{(v,k)}$.

Pour un nœud $k \in \mathcal{K}$, soit $y_{k,p}$ une variable binaire qui vaut 1 si la nouvelle position p_k du mot-clé k est p , et 0 sinon. Il s'en suit que la variation estimée $\mu_{(v,k)} - n_{(v,k)}$ de flot qui circule sur l'arc (v, k) est égale à $\sum_{p \in \mathcal{P}} y_{k,p} \Delta_{(v,k)}^p$ où

$$\Delta_{(v,k)}^p = \begin{cases} \mu_{(v,k)} - n_{(v,k)} & \text{si } p_k \neq \bar{p} \\ -n_{(v,k)} & \text{si } p_k = \bar{p} \end{cases} = \begin{cases} n_{(v,k)} [(\beta_{click})^{p - p_k^{init}} - 1] & \text{si } p_k \neq \bar{p} \\ -n_{(v,k)} & \text{si } p_k = \bar{p} \end{cases}$$

Les hypothèses A2 et A3 sur un arc (v, k) où $k \in \mathcal{K}$ imposent que toute variation de flot vers un mot-clé dont la position a été changée soit attribuée au nœud-perte ; cela dans but de modéliser ce qui est perdu à la compétition ou récupéré de la compétition. Ainsi, une partie de la différence $\mu_{(v,k)} - n_{(v,k)}$ doit être attribuée au nœud-perte l . Donc si on améliore la position de k , k attire plus de clics et le flot augmente sur (v, k) . Puisque k attire plus de clics, les compétiteurs de l'annonceur en perdent. Le flot diminue donc sur (v, l) . Et si on dégrade la position de k , c'est l'événement inverse qui se produit. Nous estimons qu'une fraction δ de la variation de flot sur (v, k) est attribuable à l . Il s'en suit que la variation de flot sur (v, l) résultant du changement est égale à $-\delta \sum_{p \in \mathcal{P}} y_{k,p} \Delta_{(v,k)}^p$.

La variation totale de flot sur (v, l) peut mener à un flot négatif. Pour éviter cela, nous la limitons à $-n_{(v, l)}$. En résumé, la variation estimée de flot sur (v, l) est égale à $-d_v$ où $d_v = \min\{n_{(v, l)}, \delta \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u, p} \Delta_{(v, u)}^p\}$.

Soit un nœud $v \in \mathcal{V}$, et $w \in \Gamma_v^+$ un de ses successeurs.

En résumé, nous fixons :

$$\begin{aligned} \mu_{(v, w)} &= n_{(v, w)} + \sum_{p \in \mathcal{P}} y_{w, p} \Delta_{(v, w)}^p && \text{si } w \in \mathcal{K} \\ \mu_{(v, w)} &= n_{(v, w)} - d_v && \text{si } w = l \\ \mu_{(v, w)} &= n_{(v, w)} && \text{si } w \notin \mathcal{K} \cup \{l\} \end{aligned}$$

Pour conserver le flot à travers le graphe, le flot total estimé μ_v traversant le nœud v est égal à la somme des flots sur les arcs sortant de v :

$$\mu_v = \sum_{w \in \Gamma_v^+} \mu_{(v, w)} = \sum_{w \in \Gamma_v^+} n_{(v, w)} - d_v + \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u, p} \Delta_{(v, u)}^p = n_v - d_v + \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u, p} \Delta_{(v, u)}^p.$$

Pour finir, nous imposons que la proportion $\alpha_{(v, w)}$ de flot traversant v vers un successeur $w \in \Gamma_v^+$ de v soit égale à sa valeur estimée. Donc, les estimations précédentes servent en fait à déterminer comment le flot sera redistribué suite au changement. On a donc :

$$x_{(v, w)} = \alpha_{(v, w)} x_v = \frac{\mu_{(v, w)}}{\mu_v} x_v \quad \text{pour tout } w \in \Gamma_v^+.$$

3.2.4 Contraintes de budget

Une autre contrainte à considérer dans le modèle est la contrainte de budget. C'est elle qui évitera de placer tous les mots-clés à la position 1. En effet, changer la position d'un mot-clé mènera au changement de son *CPC* moyen tel que présenté par Quinn (2011); Skiera et Abou Nabout (2013).

Soit $c_{k, p}$ le *CPC* moyen estimé du mot-clé $k \in \mathcal{K}$ quand il est placé à la position $p \in \mathcal{P}$. Si l'annonce n'est pas affichée, alors elle ne coûte rien : on fixe donc $c_{k, p} = 0$ quand $p = \bar{p}$. Et on utilise la formule de Quinn (2011) quand k est placé à la position $p \in \mathcal{P} \setminus \{\bar{p}\}$. Plus précisément, lorsque $p \in \mathcal{P} \setminus \{\bar{p}\}$, on fixe :

$$c_{k, p} = cp c_k (\beta_{cpc})^{p - p_k^{init}}.$$

Pour que la solution soit réalisable, la combinaison de changements des *CPCs* moyens et des flots doit respecter le budget :

$$\sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} c_{k,p} y_{k,p} + \sum_{b \in \mathcal{B}} c_p c_b x_b \leq B.$$

Certaines campagnes ont aussi un budget dédié à l'affichage. Une autre contrainte doit donc être ajoutée :

$$\sum_{b \in \mathcal{B}} c_p c_b x_b \leq B_{da}.$$

3.2.5 Modèle proposé

Le but est de maximiser la performance de la campagne. Selon le besoin de l'annonceur, la mesure de performance peut changer. Pour illustrer notre modèle, nous avons décidé de maximiser le profit de l'annonceur mais nous aurions pu choisir une autre mesure. Le modèle résultant est comme suit :

$$\max \sum_{v \in \mathcal{C}} r_v x_v - \left(\sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} c_p c_b x_b \right) \quad (3.1)$$

$$\text{s.à.} \quad \sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} c_p c_b x_b \leq B \quad (3.2)$$

$$\sum_{b \in \mathcal{B}} c_p c_b x_b \leq B_{da} \quad (3.3)$$

$$d_v = \min\{n_{(v,l)}, \delta \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u,p} \Delta_{(v,u)}^p\} \quad \forall v \in \mathcal{V} \quad (3.4)$$

$$\alpha_{(v,w)} = \frac{n_{(v,w)} + \sum_{p \in \mathcal{P}} y_{w,p} \Delta_{(v,w)}^p}{n_v - d_v + \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u,p} \Delta_{(v,u)}^p} \quad \forall v \in \mathcal{V}, \forall w \in \Gamma_v^+ \cap \mathcal{K} \quad (3.5)$$

$$\alpha_{(v,l)} = \frac{n_{(v,l)} - d_v}{n_v - d_v + \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u,p} \Delta_{(v,u)}^p} \quad \forall v \in \mathcal{V} \quad (3.6)$$

$$\alpha_{(v,w)} = \frac{n_{(v,w)}}{n_v - d_v + \sum_{u \in \Gamma_v^+ \cap \mathcal{K}} \sum_{p \in \mathcal{P}} y_{u,p} \Delta_{(v,u)}^p} \quad \forall v \in \mathcal{V}, \forall w \in \Gamma_v^+ \setminus (\mathcal{K} \cup \{l\}) \quad (3.7)$$

$$\sum_{p \in \mathcal{P}} y_{k,p} = 1 \quad \forall k \in \mathcal{K} \quad (3.8)$$

$$x_s = n_s \quad (3.9)$$

$$x_v = \sum_{u \in \Gamma_v^-} x_{(u,v)} \quad \forall v \in \mathcal{V} \setminus \{s\} \quad (3.10)$$

$$x_{(v,w)} = \alpha_{(v,w)} x_v \quad \forall (v,w) \in \mathcal{A} \quad (3.11)$$

$$x_v \geq 0 \quad \forall v \in \mathcal{V} \quad (3.12)$$

$$x_{(v,w)} \geq 0 \quad \forall (v,w) \in \mathcal{A} \quad (3.13)$$

$$y_{k,p} \in \{0, 1\} \quad \forall k \in \mathcal{K}, \forall p \in \mathcal{P} \quad (3.14)$$

Les contraintes (3.10) et (3.11) sont des contraintes de conservation de flot. La contrainte (3.9) a été ajoutée pour respecter l'hypothèse A1.

La particularité de ce modèle est qu'il devient linéaire quand les positions p_k de chaque mot-clé k sont connues. En effet, connaître p_k implique que $y_{k,p}$ est égale à 1 si et seulement si $p = p_k$. Cela veut dire que les valeurs de d_v et $\alpha_{(v,w)}$ peuvent être fixées. La valeur du reste des variables x_v et $x_{(v,w)}$ peuvent être obtenues en déterminant la solution au système d'équations (3.9)-(3.13). Nous pouvons alors déterminer si cette solution respecte les contraintes de budget (3.2) et (3.3), et calculer sa valeur selon l'objectif (3.1).

3.3 Un exemple

Dans la figure 3.3 est illustré l'impact du changement de la position d'un mot-clé sur le flot du graphe. Le nœud-source s , le nœud-conversion c , le nœud-perte l , et le puits t sont représentés en gris clair. Les deux nœud-mots-clés $k1$ et $k2$ sont en gris foncé. Les autres nœuds sont une requête (q), une bannière (b), trois nœuds ($w1, w2, w3$) représentant les pages du site du client munies d'un code de suivi du trafic, et deux conversions ($c1, c2$). Les nombres dans les boîtes sous les noms des nœuds sont le flot traversant chaque nœud. Les nombres sur les arcs correspondent à leur flot. Les arcs ayant un flot nul ne sont pas représentés. Le graphe au-dessus correspond au flot initial et celui en-dessous montre l'impact d'attribuer une position $\bar{p} = 11$ au mot-clé $k2$, sans changer la position du mot-clé $k1$. Il s'agit donc d'un cas à deux requêtes : une pour un changement de position et l'autre pour le maintien de la position courante.

Le nouveau flot est obtenu comme suit.

Supposons que $\delta = \frac{1}{4}$.

Tout d'abord, nous fixons $y_{k1, p_{k1}^{init}} = y_{k2, \bar{p}} = 1$, tandis que tous les autres $y_{k,p}$ sont à 0. Toutes les proportions $\alpha_{(v,w)}$ demeurent inchangées, à l'exception de celles avec $v = w2$. Nous avons

donc :

$$\begin{aligned}
\Delta_{(w2,k2)}^{\bar{p}} &= -n_{(w2,k2)} = -1 \\
\Delta_{(w2,k1)}^{p_{k1}^{init}} &= 0 \\
d_{w2} &= \min\{n_{(w2,l)}, \delta(\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}})\} = \min\{0, \frac{1}{4}(-1)\} = -\frac{1}{4} \\
\alpha_{(w2,k2)} &= \frac{n_{(w2,k2)} + \Delta_{(w2,k2)}^{\bar{p}}}{n_{w2} - d_v + (\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}})} = \frac{1 - 1}{4 + \frac{1}{4} + (-1)} = 0 \\
\alpha_{(w2,k1)} &= \frac{n_{(w2,k1)} + \Delta_{(w2,k1)}^{p_{k1}^{init}}}{n_{w2} - d_v + (\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}})} = \frac{2 + 0}{4 + \frac{1}{4} + (-1)} = \frac{8}{13} \\
\alpha_{(w2,w3)} &= \frac{n_{(w2,w3)}}{n_{w2} - d_v + (\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}})} = \frac{1}{4 + \frac{1}{4} + (-1)} = \frac{4}{13} \\
\alpha_{(w2,l)} &= \frac{n_{(w2,l)} - \delta\left(\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}}\right)}{n_{w2} - d_v + (\Delta_{(w2,k1)}^{p_{k1}^{init}} + \Delta_{(w2,k2)}^{\bar{p}})} = \frac{0 - \frac{1}{4}(-1)}{4 + \frac{1}{4} + (-1)} = \frac{1}{13}.
\end{aligned}$$

Le nouveau flot est alors obtenu en trouvant la solution satisfaisant (3.9)-(3.13), avec les nouvelles valeurs $\alpha_{w,x}$ pour $x = k1, k2, w3, l$.

Dans le but de faciliter la vérification qu'il y a bien conservation de flot à tous les nœuds, tous les nombres en bas de la figure 3.3 ont un dénominateur commun 507. Par exemple, le seul arc entrant $w2$ est $(s, w2)$, ce qui veut dire que le flot traversant $w2$ est égal à $\frac{2028}{507} = 4$. Il est distribué à ses successeurs comme suit :

- le montant envoyé à l est $\frac{156}{507}$, ce qui correspond à une proportion de $\frac{156}{2028} = \frac{1}{13} = \alpha_{(w2,l)}$;
- le montant envoyé à $w3$ est $\frac{624}{507}$, ce qui correspond à une proportion de $\frac{624}{2028} = \frac{4}{13} = \alpha_{(w2,w3)}$;
- le montant envoyé à $k1$ est $\frac{2028-156-624}{507} = \frac{1248}{507}$ ce qui correspond à une proportion de $\frac{1248}{2028} = \frac{8}{13} = \alpha_{(w2,k1)}$.

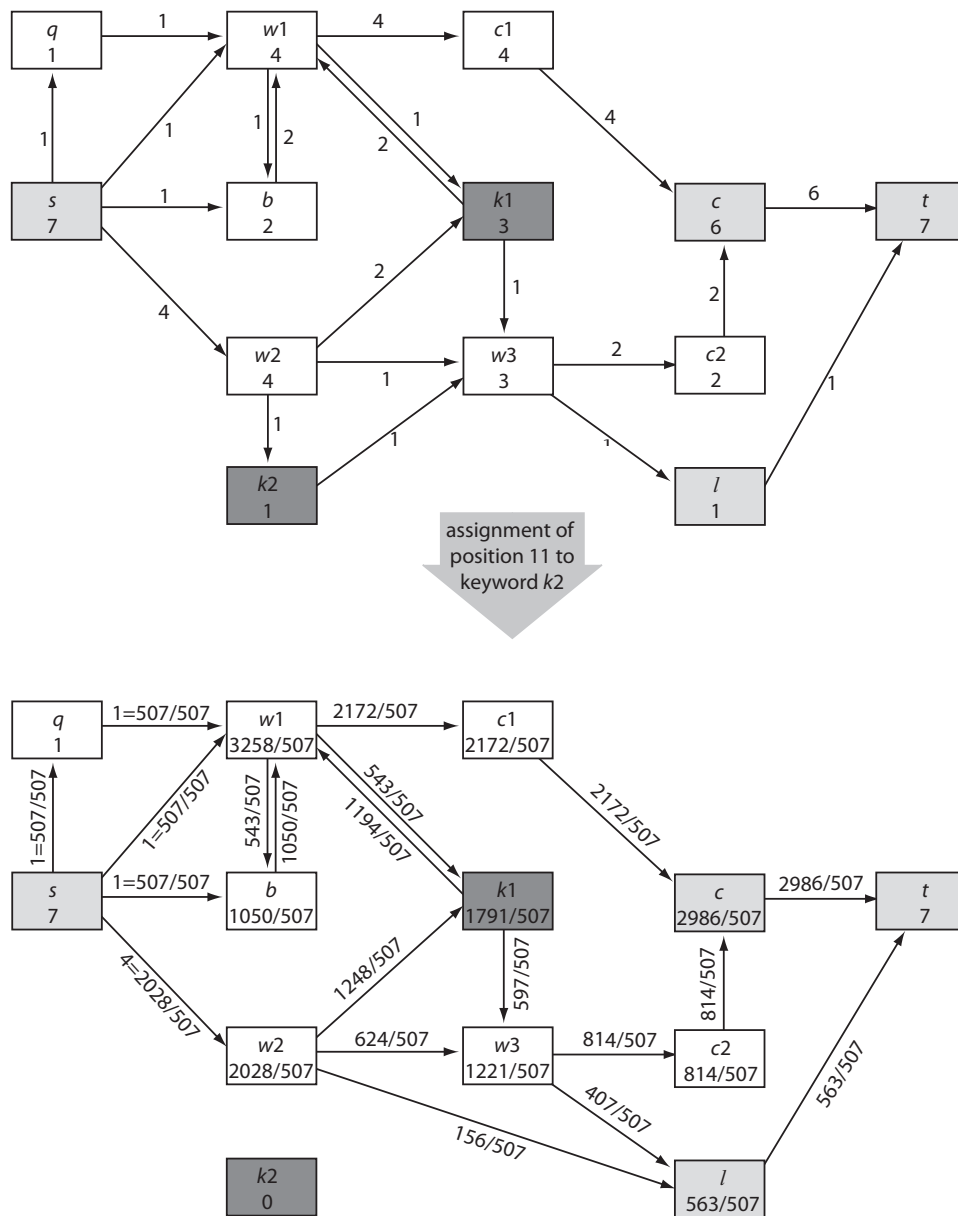


Figure 3.3 Illustration de l'assignation de la position \bar{p} à un mot-clé

3.4 Autres fonctions-objectifs

Tel que mentionné précédemment, il existe plus d'une mesure de performance d'une campagne. Bosomworth (2016) et Edgecomb classent quelques indices de performance selon leur type (ce qui est mesuré) et le but de la campagne (ce que l'annonceur cherche à accomplir). Cette classification est présentée sous la forme d'une matrice avec un but par ligne et un type par colonne. Ils ont considéré trois types de mesure :

- commercial : mesures d'affaire qui reflètent les notes de satisfaction telles que les parts d'audience et les ventes ;
- performance de contenu ou tactique : mesures liées directement à l'annonce telles que les vues, les partages ou les clics ;
- marque : mesures qui reflètent à quel point la marque est connue du public telles que le nombre de fois que le nom de la marque est mentionné au quotidien comparativement à la compétition.

Les quatre buts considérés sont définis comme suit :

- atteindre : le but est de faire connaître la marque par autant d'utilisateurs que possible ;
- agir : le but est de faire interagir les utilisateurs sur le site de l'annonceur ou sur sa page dans les réseaux sociaux ;
- convertir : le but est de mener les utilisateurs à des conversions ;
- engager : le but est de bâtir des relations à long terme avec les utilisateurs.

Avec les données disponibles, seules quelques mesures présentées dans Brady (2016) peuvent être utilisées. Nous nous concentrerons donc sur les mesures suivantes. Selon nous, elles tiennent compte des problématiques les plus importantes des annonceurs :

- Profit : une mesure commerciale qui trouve le meilleur compromis entre revenus et coûts ;

$$\max \sum_{v \in \mathcal{C}} x_v r_v - \left(\sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} x_b c_b \right)$$

- Revenu : une mesure commerciale pour ceux qui veulent générer des revenus à travers une campagne de conversions ;

$$\max \sum_{v \in \mathcal{C}} x_v r_v$$

- Nombre de visites sur le site de l'annonceur : une mesure de marque pour une campagne d'atteinte et, puisque nous supposons que le nombre total d'utilisateurs ne changera pas (hypothèse A1), on peut également la considérer pour une campagne d'engagement.

Pour nos expériences, nous maximisons le nombre de clics qui mènent à toute page du site de l'annonceur mais il est possible de choisir tout sous-ensemble d'arcs ou de nœuds. Par exemple, maximiser le flot sur les arcs allant des annonces vers le site de l'annonce serait un autre moyen de maximiser le taux de clic ;

$$\max \sum_{v \in \mathcal{W}} x_v$$

- Nombre de conversions : une mesure de marque pour une campagne de conversion et une mesure de performance de contenu d'une campagne d'engagement.

$$\max \sum_{v \in \mathcal{C}} x_v$$

CHAPITRE 4 ALGORITHMES

Pour résoudre le modèle présenté au chapitre précédent, cinq algorithmes ont été développés et combinés à au moins une des quatre méthodes de résolution de systèmes d'équations explorées. Dans ce chapitre, nous présenterons ces différents algorithmes ainsi que les méthodes. Initialement, le solveur CPLEX était utilisé pour résoudre le modèle. Toutefois, il n'arrivait pas à résoudre ce problème non-linéaire. Une heuristique ont donc été employée.

Tel que souligné dans le chapitre précédent, une fois les positions fixées, le problème devient linéaire et se limite à la résolution d'un système d'équations. L'heuristique a pour but de choisir les positions à assigner aux mots-clés alors que CPLEX était utilisé pour résoudre le système d'équations.

Pour la suite, toute solution S est un ensemble de positions assignées à des mots-clés et peut donc être définie par l'expression suivante :

$$S = \left\{ y_{k,p} \mid \sum_{p \in \mathcal{P}} y_{k,p} = 1 \quad \forall k \in \mathcal{K} \text{ et tel que } y_{k,p} \in \{0, 1\} \right\}. \quad (4.1)$$

Pour toute solution S , nous considérerons les notations suivantes :

- \mathcal{N}_S^1 le premier voisinage de S :

$$\mathcal{N}_S^1 = \mathcal{N}_S^{1-} \cup \mathcal{N}_S^{1+} \quad (4.2)$$

Où \mathcal{N}_S^{1-} est le voisinage de S obtenu en améliorant (i.e. en diminuant) la position d'un mot-clé d'une unité et \mathcal{N}_S^{1+} est le voisinage de S obtenu en détériorant (i.e. en augmentant) la position d'un mot-clé d'une unité.

- \hat{k} le mot-clé dont la position a été changée de la solution S à la meilleure solution dans \mathcal{N}_S^1
- d la variation de position de \hat{k} donc, par définition, $d \in \{-1, +1\}$
- $\mathcal{N}_S^2(\hat{k}, p)$ le second voisinage de la solution S obtenu en amplifiant ou non le changement de position de \hat{k} . Par exemple, si la position de \hat{k} est passée de 3 à 4, alors dans $\mathcal{N}_S^2(\hat{k}, p)$ sont regroupées les solutions où \hat{k} est en position 4 et plus (jusqu'à \bar{p}) et tous les autres mots-clés gardent les mêmes positions que celles qu'ils ont dans S .

Une partie du contenu provient de l'article par Azeuli et al. (2017).

4.1 Assignment des positions

4.1.1 Recherche taboue

La recherche taboue est une méthode de recherche locale qui visite un espace de recherche S en se déplaçant étape par étape à partir d'une solution $s \in S$ vers une solution voisine $s' \in N(s)$, où $N(s)$ est un sous-ensemble de S appelé le voisinage de s . Une liste taboue T interdit temporairement certains mouvements qui ramèneraient la recherche à une solution visitée récemment. La recherche taboue a été introduite dans Glover (1986). Pour plus de détails, le lecteur peut se référer à Glover et Laguna (1997). Le pseudocode suivant illustre le fonctionnement général du tabou pour un problème de minimisation :

Algorithme 3 : Tabou

Entrées : Problème d'optimisation

Sorties : s^* meilleure solution trouvée

- 1: Générer une solution initiale $s \in S$
 - 2: Fixer $T \leftarrow \emptyset$ et $s^* \leftarrow s$;
 - 3: **Tant que** les critères d'arrêt ne sont pas respectés **Répéter**
 - 4: Trouver la solution $s' \in N(s)$ avec la plus petite valeur $f(s')$ telle que le déplacement de s à s' ne soit pas dans T ou $f(s') < f(s^*)$;
 - 5: **Si** $f(s') < f(s^*)$ **alors**
 - 6: Fixer $s^* \leftarrow s'$
 - 7: **Fin Si**
 - 8: Fixer $s \leftarrow s'$ et mettre à jour T ;
 - 9: **Fin Tant que**
-

Le tabou a été envisagé ici car il est conçu pour explorer une grande variété de solutions et éviter de rester bloquer sur un optimal local.

Variantes de l'algorithme tabou infructueuses

Plus loin, sont présentées quatre variantes de l'algorithme tabou qui a été retenu mais avant d'en arriver là, quelques algorithmes, également basés sur le tabou, ont été testés et se sont avérés infructueux.

Tentative 1 Dans une première tentative, la position d'un mot-clé, choisi dans un ordre séquentiel, était modifiée d'une unité selon le déplacement qui améliorerait le plus le profit. Son pseudocode se présente comme suit :

 Algorithme 4 : Tentative 1

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

```

1:  $S \leftarrow$  solution courante
2:  $best \leftarrow S$ 
3: Répéter
4:   Pour tout mot-clé  $k \in \mathcal{K}$  Faire
5:      $S' \leftarrow S$  avec  $k$  à la position  $p_k - 1$ 
6:      $S'' \leftarrow S$  avec  $k$  à la position  $p_k + 1$ 
7:      $S \leftarrow$  la meilleure solution entre  $S'$  et  $S''$  qui respecte le budget et n'est pas taboue
      (à moins d'être meilleure que best)
8:     Si  $S$  a été détériorée alors
9:       Faire le déplacement inverse devient tabou
10:    Fin Si
11:    Si  $S$  est meilleure que best et respecte le budget alors
12:       $best \leftarrow S$ 
13:    Fin Si
14:  Fin Pour
15: Jusqu'à ce que critères d'arrêt respectés

```

À partir de cet algorithme, une variante a été explorée dans le but de réduire le temps de résolution.

Tentative 2 Dans cette variante, une fois la direction du déplacement choisie, la meilleure position dans cette direction qui respectait le budget était déterminée et assignée à chaque mot-clé. Son pseudocode se présente comme suit :

 Algorithme 5 : Tentative 2

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

```

1:  $S \leftarrow$  solution courante
2:  $best \leftarrow S$ 
3: Répéter

```

```

4:   Pour tout mot-clé  $k \in \mathcal{K}$  Faire
5:      $S' \leftarrow S$  avec  $k$  à la position  $p_k - 1$ 
6:      $S'' \leftarrow S$  avec  $k$  à la position  $p_k + 1$ 
7:      $S \leftarrow$  la meilleure solution entre  $S'$  et  $S''$  qui respecte le budget et n'est pas taboue
      (à moins d'être meilleure que  $best$ )
8:     Si  $S$  a été détériorée alors
9:       Faire le déplacement inverse devient tabou
10:    Sinon
11:       $d \leftarrow$  la variation de position de  $k$   $\triangleright d \in \{-1, 1\}$ 
12:       $p \leftarrow p_k$ 
13:      Tant que  $p$  n'est pas une position extrême Répéter
14:         $S' \leftarrow S$  avec  $k$  à la position  $p + d$ 
15:        Si  $S'$  est meilleure que  $S$  et respecte le budget alors
16:           $S \leftarrow S'$ 
17:        Fin Si
18:      Fin Tant que
19:    Fin Si
20:    Si  $S$  est meilleure que  $best$  et respecte le budget alors
21:       $best \leftarrow S$ 
22:    Fin Si
23:  Fin Pour
24: Jusqu'à ce que critères d'arrêt respectés

```

Avec cette variante, il a été constaté qu'une fois le budget épuisé, si les mots-clés n'avaient pas tous été réévalués, l'algorithme était "paralysé" et, faute de budget, ne pouvait souvent pas mettre les mots-clés restants à leur meilleure position. Cette observation a mené à une troisième tentative et, plus tard, à un algorithme différent.

Tentative 3 Dans cette tentative, pour éviter la paralysie, il est devenu permis d'explorer des solutions dépassant le budget mais elles sont pénalisées. Son pseudocode se présente comme suit :

Algorithme 6 : Tentative 3

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : $best$ meilleure solution trouvée qui respecte le budget

- 1: $S \leftarrow$ solution courante
 - 2: $best \leftarrow S$
 - 3: **Répéter**
-

```

4:   Pour tout mot-clé  $k \in \mathcal{K}$  Faire
5:      $S' \leftarrow S$  avec  $k$  à la position  $p_k - 1$ 
6:      $S'' \leftarrow S$  avec  $k$  à la position  $p_k + 1$ 
7:      $S \leftarrow$  la meilleure solution entre  $S'$  et  $S''$  qui n'est pas taboue (à moins d'être
      meilleure que  $best$ )
8:     Si  $S$  a été détériorée alors
9:       Faire le déplacement inverse devient tabou
10:    Sinon
11:       $d \leftarrow$  la variation de position de  $k$   $\triangleright d \in \{-1, 1\}$ 
12:       $p \leftarrow p_k$ 
13:      Tant que  $p$  n'est pas une position extrême Répéter
14:         $S' \leftarrow S$  avec  $k$  à la position  $p + d$ 
15:        Si  $S'$  est meilleure que  $S$  alors
16:           $S \leftarrow S'$ 
17:        Fin Si
18:      Fin Tant que
19:    Fin Si
20:    Si  $S$  est meilleure que  $best$  et respecte le budget alors
21:       $best \leftarrow S$ 
22:    Sinon Si  $S$  dépasse le budget alors
23:      Appliquer une pénalité à la valeur de  $S$ 
24:    Fin Si
25:  Fin Pour
26:  Si les 10 dernières solutions dépassaient le budget alors
27:    Doubler la pénalité
28:  Sinon Si les 10 dernières solutions respectaient le budget alors
29:    Réduire la pénalité de moitié
30:  Fin Si
31: Jusqu'à ce que critères d'arrêt respectés

```

Malheureusement, avec cette expansion de l'espace de solutions, il arrive que l'algorithme prennent plus de temps à trouver la même solution que son prédécesseur.

Algorithme tabou retenu

Comparativement à l'algorithme précédent, celui-ci cherche à déterminer le meilleur ordre dans lequel modifier les positions des mots-clés. Ainsi, les mots-clés dont le changement de position a le meilleur impact sur l'objectif verra sa position changer en premier. Quatre variantes de cette approche ont été développées dans le cadre de cette thèse.

Elles font usage des fonctions et procédures suivantes :

- **TABULIST**(S) met à jour la liste taboue \mathcal{L} . Quand un mot-clé est ajouté à \mathcal{L} , sa position ne peut pas être changée pour les n itérations suivantes.
- **BESTOF**(\mathcal{N}_S) retourne la solution dans un voisinage donné \mathcal{N}_S de la solution S avec la meilleure valeur. Elle met également à jour *best*, la meilleure solution trouvée jusqu'à là qui respecte la contrainte (3.2), et *globalBest*, la meilleure solution trouvée qui ne respecte pas nécessairement la contrainte. À moins que la solution sélectionnée ne soit meilleure que *globalBest* ou *best*, elle ne doit pas être obtenue en faisant un déplacement interdit par \mathcal{L} .
- **BUDGETBESTOF**(\mathcal{N}_S) fait la même chose que **BESTOF**(\mathcal{N}_S) sauf que la solution retournée doit respecter le budget et qu'il n'y a pas de *globalBest*.
- **COST**(S) retourne le coût d'une solution donnée S .
- **PENALTIES** : λ est une pénalité appliquée à la valeur d'une solution qui ne respecte pas la contrainte (3.2). Si les valeurs des solutions sélectionnées dépassent le budget pour dix itérations consécutives, la valeur de λ est doublée ; à l'inverse, si elles sont en-dessous ou égales au budget pour dix itérations consécutives, λ sera réduit de moitié comme dans Gendreau et al. (1994). La stratégie de pénalisation de solutions non réalisables a été introduite par Glover (1989).
- **z**(S) retourne la valeur d'une solution. Si l'algorithme utilise les pénalités, alors :

$$z(S) \leftarrow \sum_{v \in \mathcal{C}} r_v \cdot x_v^S - \text{COST}(S) - \lambda \cdot \max \{0; \text{COST}(S) - \text{budget}\} \quad (4.3)$$

Sinon

$$z(S) \leftarrow \sum_{v \in \mathcal{C}} r_v \cdot x_v^S - \text{COST}(S). \quad (4.4)$$

Tous les algorithmes tabous présentés dans cette thèse ont la même procédure d'initialisation :

 Algorithme 7 : Initialisation de l'algorithme

```

1: Procédure INITIALIZATION
2:    $n \leftarrow 2 \cdot \sqrt{|\mathcal{K}|}$ 
3:    $\mathcal{L} \leftarrow \emptyset$ 
4:    $S \leftarrow \{(k, p_k^{init}) \mid k \in \mathcal{K}\}$  ▷  $S$  est la solution courante
5:    $best \leftarrow S$ 
6:    $budget \leftarrow \text{COST}(S)$ 
7: Fin Procédure

```

Bien que la solution initiale utilisée ici est celle fournie par les données, elle peut aussi être toute autre solution générée aléatoirement. Toutefois, les deux premières variantes nécessitent que la solution initiale respecte le budget.

Première variante (Tabou₁) L'adaptation suivante de la recherche taboue a été développée pour trouver une bonne solution au modèle proposé. Une solution est définie par une assignation de position à chaque mot-clé $k \in \mathcal{K}$. En d'autres mots, une solution fixe la valeur de chaque variable $y_{k,p}$. Tel que mentionné plus haut, le flot F_s associé à chaque solution s peut être déterminé en calculant les valeurs des variables de flot qui satisfont les équations (3.9)-(3.13). L'espace de recherche S contient toutes les solutions s telles que F_s satisfait la contrainte de budget (3.2). L'objectif (3.1) est utilisé pour évaluer chaque solution $s \in S$. Le voisinage $N(s)$ de s contient toutes les solutions $s' \in S$ qui peuvent être obtenues de s en ajoutant une unité (si $p_k < \bar{p}$) ou en soustrayant une unité (si $p_k > 1$) à la position p_k d'un seul mot-clé $k \in \mathcal{K}$. Quand on passe d'une solution s à un voisin $s' \in N(s)$, le mot-clé k auquel une nouvelle position est assigné entre dans la liste taboue T et il est interdit d'en changer la position de nouveau pour $2\sqrt{|\mathcal{K}|}$ itérations, à moins que cette modification améliore la meilleure solution s^* (i.e. *best*) trouvée jusqu'ici. Le pseudocode de cet algorithme se présente comme suit :

 Algorithme 8 : Tabou₁

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

- 1: INITIALIZATION
 - 2: **Répéter**
 - 3: Déterminer \mathcal{N}_S^1
 - 4: $S \leftarrow \text{BUDGETBESTOF}(\mathcal{N}_S^1)$
 - 5: TABULIST
 - 6: **Jusqu'à ce que** les critères d'arrêt soient respectés
-

Deuxième variante (Tabou₂) Tabou₂ est similaire à Tabou₁, sauf que la modification de la position du mot-clé k n'est pas limitée à une unité lorsqu'on passe de la solution s à $s' \in N(s)$. En effet, après avoir trouvé la meilleure solution s_1 dans \mathcal{N}_s^1 , on l'améliore en cherchant la meilleure solution s_2 dans $\mathcal{N}_{s_1}^2$ (i.e. en amplifiant le déplacement de s à s_1). Le pseudocode de cet algorithme se présente comme suit :

 Algorithme 9 : Tabou₂

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

- 1: INITIALIZATION
 - 2: **Répéter**
 - 3: Déterminer \mathcal{N}_S^1
 - 4: Extraire \hat{k} et d de $\text{BUDGETBESTOF}(\mathcal{N}_S^1)$
 - 5: Déterminer $\mathcal{N}_S^2(\hat{k}, d)$
 - 6: $S \leftarrow \text{BUDGETBESTOF}(\mathcal{N}_S^2(\hat{k}, d))$
 - 7: TABULIST
 - 8: **Jusqu'à ce que** les critères d'arrêt soient respectés
-

Troisième variante (Tabou₃) Tabou₃ est similaire à Tabou₁, sauf que nous acceptons de visiter les solutions qui ne respectent pas la contrainte de budget (3.2). Cela veut dire que l'espace de recherche S contient toutes les assignations possibles de positions aux mots-clés. Une composante peut être ajoutée à la fonction-objectif de sorte que les solutions qui ne respectent pas le budget soient pénalisées. Plus précisément, la valeur de la solution $s \in S$

est égale à

$$\sum_{v \in \mathcal{C}} r_v x_v - \left(\sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} c_p c_b x_b \right) - \lambda_1 \left(\max\{0, \sum_{k \in \mathcal{K}} x_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} c_p c_b x_b - B\} \right) - \lambda_2 \left(\max\{0, \sum_{b \in \mathcal{B}} c_p c_b x_b - B_{da}\} \right)$$

Où λ_1 et λ_2 sont des facteurs de pénalité initialisés à 1. Le pseudocode de cet algorithme se présente comme suit :

Algorithme 10 : Tabou₃

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

- 1: INITIALIZATION
 - 2: $\lambda \leftarrow 1$
 - 3: *globalBest* \leftarrow *best*
 - 4: **Répéter**
 - 5: Déterminer \mathcal{N}_S^1
 - 6: $S \leftarrow \text{BESTOF}(\mathcal{N}_S^1)$
 - 7: TABULIST
 - 8: PENALTIES
 - 9: **Jusqu'à ce que** les critères d'arrêt soient respectés
-

Quatrième variante (Tabou₄) Tabou₄ est l'algorithme obtenu de Tabou₂ exactement comme Tabou₃ a été obtenu de Tabou₁ : les solutions non réalisables peuvent être visitées et un mot-clé peut modifier sa position par plus d'une unité lorsque passant d'une solution s à un voisin $s' \in N(s)$. Le pseudocode de cet algorithme se présente comme suit :

Algorithme 11 : Tabou₄

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

- 1: INITIALIZATION
 - 2: $\lambda \leftarrow 1$
 - 3: *globalBest* \leftarrow *best*
-

```

4: Répéter
5:   Déterminer  $\mathcal{N}_S^1$ 
6:   Extraire  $\hat{k}$  et  $d$  de  $\text{BUDGETBESTOF}(\mathcal{N}_S^1)$ 
7:   Déterminer  $\mathcal{N}_S^2(\hat{k}, d)$ 
8:    $S \leftarrow \text{BUDGETBESTOF}(\mathcal{N}_S^2(\hat{k}, d))$ 
9:   TABULIST
10:  PENALTIES
11: Jusqu'à ce que les critères d'arrêt soient respectés

```

4.1.2 Algorithme glouton

Tel qu'il sera possible de constater dans le chapitre suivant, bien que performant, les algorithmes tabous sont excessivement lents. Par conséquent, un glouton a également été implémenté. Il fonctionne comme suit. Étant donnée une solution s , nous déterminons pour chaque mot-clé k la meilleure valeur qui peut être obtenue en augmentant, diminuant ou en maintenant sa position courante, les modifications n'étant pas limitées à une unité. La valeur de la modification est mesurée en utilisant la même fonction-objectif comme dans Tabou₂ et Tabou₄, ce qui veut dire que les dépassements de budget sont permis. Si la modification de la position du mot-clé mène à une solution réalisable (donc qui satisfait les contraintes de budget) de meilleure valeur que celle de la meilleure solution trouvée s^* , alors s^* est mis à jour en conséquence. Nous appliquons alors ces modifications simultanément et considérons la solution en résultant comme la nouvelle solution courante s pour la prochaine itération. Le facteur de pénalité λ est mis à jour toutes les 10 itérations comme dans Tabou₃ et Tabou₄. Le processus s'arrête quand s ne change pas d'une itération à la prochaine et le résultat de l'algorithme est alors s^* . Le pseudocode de cet algorithme se présente comme suit :

Algorithme 12 : Glouton

Entrées : Graphe $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ représentant le problème d'optimisation, budget B de la campagne

Sorties : *best* meilleure solution trouvée qui respecte le budget

```

1: INITIALIZATION
2:  $\lambda \leftarrow 1$ 
3:  $globalBest \leftarrow best$ 
4: Répéter
5:   Déterminer  $\mathcal{N}_S^1$ 
6:   Trier les solutions dans  $\mathcal{N}_S^1$  de sorte que la valeur de la solution  $S_m$  soit meilleure que celle de la solution  $S_n$  si  $m < n$ ,  $\forall (s_m, s_n) \in \mathcal{N}_S^1 \times \mathcal{N}_S^1$ 

```

```

7:   Pour  $j = 1$  to  $|\mathcal{N}_S^1|$  Faire
8:      $k \leftarrow$  mot-clé dont la position a été changée de  $S$  à  $S_j$ 
9:      $p \leftarrow p_k^{S_j}$ 
10:     $d \leftarrow p - p_k^S$ 
11:    Tant que  $1 \leq p \leq \bar{p}$  Répéter
12:       $\hat{S}_j \leftarrow$  solution avec  $k$  à la position  $p$ 
13:      Si  $z(\hat{S}_j)$  est meilleur que  $z(globalBest)$  alors
14:         $globalBest \leftarrow \hat{S}_j$ 
15:        Si  $z(\hat{S}_j)$  est meilleur que  $z(best)$  ET  $COST(\hat{S}_j) \leq budget$  alors
16:           $best \leftarrow \hat{S}_j$ 
17:        Fin Si
18:      Fin Si
19:       $p \leftarrow p + d$ 
20:    Fin Tant que
21:  Fin Pour
22: Jusqu'à ce que les critères d'arrêt soient respectés

```

Le fait de ne pas réévaluer chaque changement de position avant de l'appliquer permet de gagner beaucoup de temps comparativement aux tabous. Toutefois, cette logique naïve peut amener l'algorithme à passer à côté d'une meilleure solution.

4.2 Résolution de système d'équations linéaires : Méthodes itératives

Les quatre algorithmes de recherche taboue et le glouton utilisent une procédure qui détermine le flot F_s associé à une solution s en déterminant les valeurs des variables qui satisfont (3.9)-(3.13). Nous avons comparé quatre techniques capables de résoudre ce type de système linéaire. Une possibilité est d'utiliser le solveur CPLEX. Les autres options sont d'utiliser les méthodes itératives de Jacobi ou de Gauss-Seidel telles que présentées dans Fortin (2011) ou une variante de la méthode de Gauss-Seidel que nous appellerons la méthode de Gauss-Seidel optimisée.

Ces trois méthodes supposent que tous les coefficients de la diagonale de la matrice A sont non nuls.

4.2.1 Méthode de Jacobi

À chaque itération, la méthode de Jacobi isole une des variables et en estime la valeur en utilisant les valeurs des autres variables obtenues à l'itération précédente. Soit un système de n équations à n inconnues de type $A\vec{x} = \vec{b}$ où $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ est une matrice carrée de taille n , x et b des vecteurs à n composantes, à la k^e itération, la valeur de variable x_i sera

calculée en utilisant la formule suivante :

$$x_i^k = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{k-1} \right).$$

Le pseudocode de cette méthode se présente comme suit :

Algorithme 13 : Jacobi

Entrées : Matrice des coefficients A , vecteur des membres de droite b

Sorties : Vecteur $\vec{\hat{x}}$ représentant la solution du système d'équations $A\vec{\hat{x}} = \vec{b}$

1: $\vec{x} \leftarrow$ vecteur avec le flot courant traversant chaque nœud

2: $end \leftarrow n$

3: $it \leftarrow 0$

4: **Répéter**

5: $\hat{x}_i \leftarrow \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j \right)$

6: $\vec{x} \leftarrow \vec{\hat{x}}$

7: **Jusqu'à ce que** les critères d'arrêt soient respectés

4.2.2 Méthode de Gauss-Seidel

La différence entre la méthode de Jacobi et celle de Gauss-Seidel est que cette dernière utilise les valeurs les plus à jour. Autrement dit, dans un système de n équations à n inconnues, à la k^e itération, la valeur de la variable x_i sera calculée en utilisant la formule suivante :

$$x_i^k = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} \hat{x}_j - \sum_{j=i+1}^n a_{ij} x_j \right)$$

Le pseudocode de cette méthode se présente comme suit :

Algorithme 14 : Gauss-Seidel

Entrées : Matrice des coefficients A , vecteur des membres de droite b

Sorties : Vecteur $\vec{\hat{x}}$ représentant la solution du système d'équations $A\vec{\hat{x}} = \vec{b}$

1: $\vec{x} \leftarrow$ vecteur avec le flot courant traversant chaque nœud

2: $end \leftarrow n$

3: $it \leftarrow 0$

```

4: Répéter
5:    $\hat{x}_i \leftarrow \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} \hat{x}_j - \sum_{j=i+1}^n a_{ij} x_j \right)$ 
6:    $\vec{x} \leftarrow \vec{\hat{x}}$ 
7: Jusqu'à ce que les critères d'arrêt soient respectés

```

4.2.3 Méthode de Gauss-Seidel optimisée

Puisque l'ordre des variables peut avoir un impact important sur la performance de la méthode de Gauss-Seidel, nous avons implémenté une version qui utilise l'ordre suivant. Soit $s' \in N(s)$ un voisin de s et k le mot-clé ayant différentes positions dans s et s' . Nous considérons toutes les variables liées à k (notamment x_k , $x_{(v,k)}$ où $v \in \Gamma_k^-$ et $x_{(k,v)}$ où $v \in \Gamma_k^+$). Nous considérons ensuite celles liées à ses parents et ses enfants, puis ses grands-parents et ses petits-enfants et ainsi de suite. Une itération de la méthode met à jour la valeur de chaque variable dans le même ordre (voir lignes 8-19 du pseudocode 15). Ce processus est répété jusqu'à ce que la précision désirée soit atteinte ou que cents itérations aient été complétées, selon ce qui se produit en premier. Le pseudocode de cette méthode se présente comme suit :

Algorithme 15 : Gauss-Seidel optimisée

Entrées : Matrice des coefficients A , vecteur des membres de droite b , indice m_k du vecteur représentant le mot-clé k dont la position a été changée

Sorties : Vecteur $\vec{\hat{x}}$ représentant la solution du système d'équations $A\vec{x} = \vec{b}$

```

1:  $\vec{x} \leftarrow$  vecteur avec le flot courant traversant chaque nœud
2:  $end \leftarrow \max \{m_k, |\mathcal{V}| - m_k\}$ 
3:  $it \leftarrow 0$ 
4: Répéter
5:    $\hat{x}_k \leftarrow \frac{1}{a_{m_k, m_k}} \left( b_k - \sum_{\substack{j=1 \\ j \neq m_k}}^{|\mathcal{V}|} a_{m_k, j} \cdot x_j \right)$ 
6:    $j_1 \leftarrow m_k$ 
7:    $j_2 \leftarrow m_k + 1$ 
8:   Pour  $i = 1$  to  $end$  Faire
9:     Si  $j_1 > 1$  alors
10:        $j_1 \leftarrow m_k - i$ 
11:        $\Omega_1 \leftarrow \{1, \dots, j_1 - 1\} \cup \{j_2, \dots, |\mathcal{V}|\}$ 
12:        $\hat{x}_{j_1} \leftarrow \frac{1}{a_{j_1, j_1}} \left( b_{j_1} - \sum_{j \in \Omega_1} a_{j_1, j} \cdot x_j - \sum_{j=j_1+1}^{j_2-1} a_{j_1, j} \cdot \hat{x}_j \right)$ 
13:     Fin Si

```

```

14:      Si  $j_2 < |\mathcal{V}|$  alors
15:           $j_2 \leftarrow m_k + i$ 
16:           $\Omega_2 \leftarrow \{1, \dots, j_1 - 1\} \cup \{j_2 + 1, \dots, |\mathcal{V}|\}$ 
17:           $\hat{x}_{j_2} \leftarrow \frac{1}{a_{j_2, j_2}} \left( b_{j_2} - \sum_{j \in \Omega_2} a_{j_2, j} \cdot x_j - \sum_{j=j_1}^{j_2-1} a_{j_2, j} \cdot \hat{x}_j \right)$ 
18:      Fin Si
19:      Fin Pour
20:       $\vec{x} \leftarrow \vec{\hat{x}}$ 
21: Jusqu'à ce que les critères d'arrêt soient respectés

```

CHAPITRE 5 TESTS EXPÉRIMENTAUX

Dans ce chapitre, nous présenterons les résultats des différentes expérimentations que nous avons réalisées dans le cadre de cette thèse. Cela inclut les tests documentés dans l'article par Azeuli et al. (2017).

5.1 Première série de tests

Dans cette première série de tests, nous allons tester le modèle initial et les algorithmes. Pour cela, nous avons généré 12 instances notées G_1, \dots, G_{12} , avec des proportions données de chaque type de nœuds (mot-clé, bannières, etc.). Six d'entre elles ont 50 mots-clés, et les autres en ont 500. Des chemins aléatoires ont été générés selon les règles suivantes :

- un chemin peut commencer avec tout type de nœud à l'exception des conversions ;
- un nœud représentant une page du site de l'annonceur peut être suivi par n'importe quel élément ;
- un nœud représentant une conversion marque toujours la fin d'un chemin ;
- tout nœud, exceptions faites d'une conversion et d'une page du site de l'annonceur, est toujours suivi d'une page du site de l'annonceur ;
- un chemin ne peut se terminer que par une conversion ou une page du site de l'annonceur.

Ces règles sont déduites des données réelles fournies par le partenaire industriel. Il ne nous a malheureusement pas été possible de faire des tests avec ces données car nous n'avons pas pu obtenir l'approbation d'un de leurs clients pour une période suffisamment longue.

Le *CPC* moyen cpc_b de toute bannière $b \in \mathcal{B}$, le *CPC* moyen initial cpc_k de tout mot-clé $k \in \mathcal{K}$ et le revenu r_c de toute conversion ont été choisis aléatoirement dans $[0.5, 10]$, $[0.01, 5]$, et $[10, 100]$, respectivement. Ces valeurs ont été déduites des données fournies par notre partenaire industriel et il a confirmé que ces valeurs sont acceptables. Nous avons fixé $\beta_{click} = 0.843$ et $\beta_{cpc} = 0.613$, tel que suggéré dans Quinn (2011). Aussi, la valeur $\frac{1}{20}$ a été arbitrairement choisie pour δ . Cette valeur semblait raisonnable pour notre hypothèse. Pour chaque mot-clé, nous avons choisi une position aléatoire dans $\{1, \dots, 10\}$. Sachant que n_v est le flot initial traversant le nœud v , nous avons donc une solution s_i pour tout graphe G_i avec un coût $(\sum_{k \in \mathcal{K}} n_k \sum_{p \in \mathcal{P}} y_{k,p} c_{k,p} + \sum_{b \in \mathcal{B}} cpc_b n_b)$, un revenu $(\sum_{v \in \mathcal{C}} r_v n_v)$ et un profit (revenu-coût), qui sont donnés dans le tableau 5.1 pour comparaison avec les résultats produits par les algorithmes proposés. Le budget B d'une campagne est fixé comme égal au

coût de la solution s_i , alors que le budget B_{da} pour une campagne d’affichages est fixé comme égal à $\sum_{b \in \mathcal{B}} cpc_b n_b$. Ce sont des hypothèses raisonnables car les annonceurs ont tendance à utiliser l’intégrité de leur budget.

Les caractéristiques des douze instances sont résumées dans le tableau 5.1. Nous y indiquons :

- les identifiants des graphes ;
- le nombre total de nœuds, le nombre de nœuds de chaque type et le nombre total d’arcs ;
- le nombre total de chemins, le nombre de chemins qui ont mené à une conversion ($\sum_{v \in \mathcal{C}} n_v$) ou à une perte (n_l), et le nombre total de visites sur toute page du site de l’annonceur ($\sum_{v \in \mathcal{W}} n_v$) ;
- les coût, revenu et profit de tout s_i .

Tableau 5.1 Caractéristiques des instances

Graphe	$ \mathcal{V} $	$ \mathcal{K} $	$ \mathcal{B} $	$ \mathcal{C} $	$ \mathcal{W} $	$ \mathcal{Q} $	$ \mathcal{A} $	chemins	conv.	pertes	visites	coûts	revenu	profit
G_1	545	50	104	8	95	284	2190	1000	8	992	1437	1729.91	447.20	-1282.71
G_2	292	50	25	23	77	113	1138	500	23	477	755	516.28	1271.30	755.02
G_3	146	50	27	1	32	32	456	200	1	199	308	571.95	58.60	-513.35
G_4	505	50	129	57	91	174	2026	1000	57	943	1352	2532.21	3076.50	544.29
G_5	279	50	23	15	88	99	936	500	15	485	671	539.74	660.30	120.56
G_6	144	50	26	14	45	5	376	200	14	186	280	577.00	827.60	250.60
G_7	5413	500	1689	663	1879	678	18449	10000	663	9337	12865	30791.61	35799.90	5008.29
G_8	2921	500	663	253	854	647	9859	5000	253	4747	6809	12280.91	13954.60	1673.69
G_9	1416	500	354	46	289	223	4631	2000	46	1954	3076	7613.96	2303.20	-5310.76
G_{10}	5005	500	479	506	1188	2328	19347	10000	506	9494	13306	9643.65	29245.70	19602.05
G_{11}	2754	500	752	142	271	1085	11452	5000	142	4858	7336	14532.41	8248.30	-6284.11
G_{12}	1404	500	171	150	91	488	5090	2000	150	1850	3158	4554.96	8665.80	4110.84

Cinq algorithmes (quatre tabous et un glouton) et quatre techniques pour résoudre le système d’équations linéaires (3.9)-(3.13) ont été présentés, ce qui donne un total de 20 méthodes possibles pour résoudre le problème à l’étude.

Pour tous les graphes, nous avons exécuté trois des algorithmes avec 15 solutions de départ différentes : une avec tous les mots-clés à la première position, une avec tous les mots-clés en position $\bar{p} = 11$, et les 13 autres avec les mots-clés à des positions aléatoirement choisies dans $\{1, \dots, 10\}$. Le meilleur temps, le pire et le temps moyen de calcul (en secondes) de chaque algorithme sur les 15 exécutions sont notés \underline{t}_y^x , \bar{t}_y^x , et t_y^x , respectivement, avec $x \in \{T3, T4, Gr\}$ et $y \in \{C, J, GS, oGS\}$, où $T3, T4$ sont utilisés pour Tabou₃, Tabou₄, respectivement, Gr pour le glouton, C pour CPLEX (version 12.2), J pour Jacobi, GS pour Gauss-Seidel, et oGS pour Gauss-Seidel optimisé (avec l’ordre spécial mentionné dans la section précédente).

Les algorithmes Tabou₁ et Tabou₂, notés $T1$ et $T2$ respectivement, n'ont été testés qu'avec la deuxième solution de départ (i.e. celle avec tous les mots-clés en position \bar{p}) car ils nécessitent une solution de départ réalisable. Le profit moyen obtenu avec l'algorithme x combiné à la méthode y pour la résolution du système linéaire est noté z_y^x . Tous les tests ont été effectués sur une machine avec un processeur Intel Core i7 2.67 GHz et 8 Go de mémoire vive.

Nous indiquons dans le tableau 5.2, le meilleur temps, le pire et le temps moyen de calcul t_{oGS}^{Gr} , \bar{t}_{oGS}^{Gr} , et t_{oGS}^{Gr} obtenu avec le glouton combiné à la méthode Gauss-Seidel optimisée. Ces valeurs sont comparées à celles obtenues avec le glouton combiné aux trois autres méthodes de résolutions de systèmes linéaires. Plus précisément, pour $y \in \{C, J, GS\}$, nous indiquons les ratios $\frac{t_y^{Gr}}{t_{oGS}^{Gr}}$, $\frac{\bar{t}_y^{Gr}}{\bar{t}_{oGS}^{Gr}}$, et $\frac{t_y^{Gr}}{t_{oGS}^{Gr}}$.

Les temps donnés sont en secondes. Les enchères pour le placement des annonces sont complétées en quelques centaines de millisecondes tout au plus. Il est donc évident que pour la plupart des graphes, les temps de résolution obtenus sont trop importants. Toutefois, étant donné nos capacités limitées en programmation, il est quasi-certain qu'il est possible d'optimiser le code utilisé pour avoir des temps de résolution beaucoup plus raisonnables ; auquel cas, il est très peu probable que cela affecte les ratios présentés.

Tableau 5.2 Temps de calcul de l'algorithme glouton avec quatre différentes méthodes de résolution de systèmes d'équations linéaires

Graphe	Gauss-Seidel optimisé			CPLEX			Jacobi			Gauss-Seidel		
	t_{oGS}^{Gr}	\bar{t}_{oGS}^{Gr}	t_{oGS}^{Gr}	$\frac{t_C^{Gr}}{t_{oGS}^{Gr}}$	$\frac{\bar{t}_C^{Gr}}{\bar{t}_{oGS}^{Gr}}$	$\frac{t_C^{Gr}}{t_{oGS}^{Gr}}$	$\frac{t_J^{Gr}}{t_{oGS}^{Gr}}$	$\frac{\bar{t}_J^{Gr}}{\bar{t}_{oGS}^{Gr}}$	$\frac{t_J^{Gr}}{t_{oGS}^{Gr}}$	$\frac{t_{GS}^{Gr}}{t_{oGS}^{Gr}}$	$\frac{\bar{t}_{GS}^{Gr}}{\bar{t}_{oGS}^{Gr}}$	$\frac{t_{GS}^{Gr}}{t_{oGS}^{Gr}}$
G_1	0.70	2.71	1.65	19.03	20.99	21.23	3.80	3.34	3.74	1.00	0.90	0.98
G_2	0.31	0.76	0.53	80.78	76.67	78.26	3.88	3.92	3.95	0.96	1.00	0.99
G_3	0.08	0.23	0.16	253.29	237.32	239.22	3.29	3.53	3.52	0.96	0.99	0.99
G_4	0.85	2.30	1.39	22.02	24.90	23.22	3.34	2.90	3.35	1.00	0.94	0.99
G_5	0.27	0.61	0.41	88.85	103.11	91.08	3.10	3.27	3.29	0.95	0.99	1.00
G_6	0.08	0.17	0.12	244.99	302.01	294.40	2.91	3.43	3.29	0.89	1.02	0.98
G_7	1093.04	1997.93	1495.97	1.19	0.83	1.01	3.48	3.49	3.48	1.01	1.01	1.01
G_8	346.22	746.75	545.49	1.36	1.18	1.31	4.74	3.59	3.60	1.03	1.03	1.04
G_9	75.36	226.69	148.55	3.49	2.67	2.98	3.68	3.70	3.69	1.07	1.06	1.07
G_{10}	1027.09	2112.46	1522.72	1.86	1.88	2.03	3.28	3.32	3.31	1.02	1.04	1.02
G_{11}	317.45	737.52	525.71	1.69	1.43	1.65	3.55	3.50	3.55	1.04	1.03	1.03
G_{12}	88.60	217.78	154.81	3.59	3.21	3.30	3.62	3.55	3.59	1.07	1.04	1.07

Les méthodes Gauss-Seidel et Gauss-Seidel optimisée sont plus rapides que les méthodes Jacobi et CPLEX. Pour les graphes G_3 et G_6 , CPLEX est jusqu'à 200 fois plus lent. La méthode Jacobi est généralement 3 à 4 plus lentes que les deux versions de la méthode Gauss-Seidel. Pour de petites instances (G_i avec $1 \leq i \leq 6$), la version non optimisée de la méthode Gauss-Seidel est un peu plus rapide que la version optimisée, alors que pour les plus

grandes (G_i avec $7 \leq i \leq 12$) c'est le contraire. Cela n'est pas très surprenant car chaque changement de position n'affecte que les ancêtres et les descendants d'un nœud représentant le mot-clé dont la position vient d'être changée. Ainsi en commençant par ce nœud et en continuant avec ces ancêtres et descendants, le calcul du nouveau flot est complété plus rapidement que lorsqu'on suit un autre ordre.

Au cours des expérimentations, il a été observé que le second scénario, qui commence avec tous les mots-clés en position \bar{p} , est généralement le plus rapide et donne la borne inférieure t_y^{Gr} pour tout $y \in \{C, J, GS, oGS\}$. Pour ces raisons, les résultats qui suivent sont tous obtenus avec ce scénario et la version optimisée de la méthode Gauss-Seidel.

Dans le tableau 5.3, les temps de calcul et les profits obtenus par l'algorithme glouton sont comparés à ceux produits par les tabous. Le critère d'arrêt des tabous était que la différence en valeur absolue $|f(s') - f(s)|$ soit plus petite que 10^{-2} entre deux solutions consécutivement sélectionnées. Nous indiquons dans le tableau 5.3 le temps de calcul du glouton et les ratios $\frac{t_{oGS}^x}{t_{oGS}^{Gr}}$ pour $x \in \{T1, T2, T3, T4\}$. Le tableau 5.3 contient également le profit z_{oGS}^{Gr} obtenu avec le glouton ainsi que les différences $z_{oGS}^{Gr} - z_{oGS}^x$ où $x \in \{T1, T2, T3, T4\}$.

Les algorithmes tabous sont beaucoup plus lents que le glouton. Pour les instances à 500 mots-clés, les ratios des temps de calcul de Tabou₁ et Tabou₃ atteignent 100. Aussi, le glouton a de meilleurs profits que les tabous. Cela est dû au critère d'arrêt. En effet, en arrêtant les tabous quand la différence en valeur absolue $|f(s') - f(s)|$ est plus petite que 10^{-3} (au lieu de 10^{-2}) entre deux solutions consécutivement sélectionnées, on obtient alors de meilleurs profits. Par exemple, ce critère d'arrêt modifié permet au Tabou₂ de trouver une solution avec $z_{oGS}^{Gr} - z_{oGS}^{T2} = -3,59 \cdot 10^{-6}$, mais le temps de calcul est alors multiplié par 4. De plus, nous avons exécuté tous les tabous plus longtemps. Même s'ils parviennent aux mêmes profits que le glouton, cela leur nécessite une très grande quantité de temps pour gagner seulement quelques dollars.

Tableau 5.3 Comparaisons du glouton aux quatre tabous combinés à la méthode Gauss-Seidel optimisée

Graphe	Temps de calcul					profits				
	Glouton	Tabou ₁	Tabou ₂	Tabou ₃	Tabou ₄	Glouton	Tabou ₁	Tabou ₂	Tabou ₃	Tabou ₄
	t_{oGS}^{Gr}	$\frac{t_{oGS}^{T1}}{t_{oGS}^{Gr}}$	$\frac{t_{oGS}^{T2}}{t_{oGS}^{Gr}}$	$\frac{t_{oGS}^{T3}}{t_{oGS}^{Gr}}$	$\frac{t_{oGS}^{T4}}{t_{oGS}^{Gr}}$	z_{oGS}^{Gr}	$z_{oGS}^{Gr}-z_{oGS}^{T1}$	$z_{oGS}^{Gr}-z_{oGS}^{T2}$	$z_{oGS}^{Gr}-z_{oGS}^{T3}$	$z_{oGS}^{Gr}-z_{oGS}^{T4}$
G_1	0.70	7.72	75.26	7.78	73.71	-1137.56	0.01	0.00	0.01	0.00
G_2	0.31	45.08	44.23	44.48	44.58	916.68	0.06	0.00	0.06	0.00
G_3	0.08	9.40	5.52	9.32	5.39	-370.29	0.73	0.00	0.73	0.00
G_4	0.85	28.09	13.20	28.05	13.09	713.23	0.08	0.00	0.08	0.00
G_5	0.27	14.11	12.37	14.35	12.62	305.86	0.18	0.01	0.18	0.01
G_6	0.08	27.55	6.00	30.39	6.61	507.53	1.12	1.09	1.12	1.09
G_7	1093.04	102.13	27.05	101.85	27.04	7243.88	14.30	0.96	14.30	0.96
G_8	346.22	81.42	26.63	81.27	26.69	3742.09	20.44	0.95	20.44	0.95
G_9	75.36	54.42	24.97	53.56	25.25	-3719.84	25.22	1.24	25.22	1.24
G_{10}	1027.09	90.95	25.76	91.17	25.91	21462.73	16.78	0.98	16.78	0.98
G_{11}	317.45	57.98	22.88	58.38	23.26	-4843.32	28.54	1.97	28.54	1.97
G_{12}	88.60	59.51	29.66	59.03	29.08	5752.34	41.43	0.80	41.43	0.80

5.2 Deuxième série de tests

En prévision de la rédaction d'un second article sur les analyses d'impact, dix nouvelles instances G_{13}, \dots, G_{22} ont été générées de la même manière que les douze instances précédentes. Pour les expériences suivantes, sauf si indiqué autrement, ce sont ces nouvelles instances qui sont utilisées. Leurs caractéristiques sont résumées dans le tableau 5.4.

Tableau 5.4 Caractéristiques des instances

Graphe	V	K	B	W	C	Q	A	chemins conv. pertes visites				coûts	revenu	profit
G_{13}	378	150	52	89	22	61	1238	750	22	728	1012	1692.72	1192.00	-500.72
G_{14}	504	150	33	173	86	58	1443	1000	86	914	1240	1205.45	4694.90	3489.45
G_{15}	605	300	23	120	27	131	2000	1200	27	1173	1653	1364.99	1584.90	219.91
G_{16}	1504	300	83	562	8	547	5089	3000	8	2992	3929	2260.36	444.50	-1815.86
G_{17}	904	450	67	192	29	162	2960	1800	29	1771	2478	2834.18	1894.70	-939.48
G_{18}	2255	450	252	425	91	1033	9042	4500	91	4409	6234	6155.87	4711.60	-1444.27
G_{19}	1204	600	102	143	286	69	3870	2400	286	2114	3170	6241.31	15883.00	9641.69
G_{20}	1505	600	15	386	101	399	4951	3000	101	2899	4013	1809.94	5388.50	3578.56
G_{21}	1879	750	407	207	407	104	6884	3750	407	3343	5010	14882.64	22428.20	7545.56
G_{22}	2503	750	301	386	575	487	9117	5000	575	4425	6537	10313.22	32952.50	22639.28

Tel qu'observé dans la première série de tests, bien le tabou puisse produire de meilleurs résultats, il est trop lent et la différence n'est pas assez importante pour justifier son utilisation pour la suite. Puisque le glouton combiné à la version optimisée de la méthode Gauss-Seidel était généralement le plus rapide dans la première série de tests, c'est uniquement cette combinaison que nous utiliserons dans cette deuxième série de tests, sauf pour la première

vague de la comparaison des fonctions-objectifs où elle est comparée au Tabou₂. De plus, pour la même raison, à moins d'indication contraire, le deuxième scénario pour la solution de départ (tous les mots-clés en position $\bar{p} = 11$) est utilisé. La valeur de δ étant purement hypothétique, il est très peu probable que cela affecte les résultats présentés à la section 5.1. Aussi, pour la suite des tests, elle a été arbitrairement fixée à 0.01, sauf si indication contraire. Tous les autres paramètres sont les mêmes que ceux de la première série de tests. Tous les tests suivants ont été exécutés sur une machine avec un processeur Intel Core i7-4578 à 3.00Ghz et une mémoire vive de 16Go.

Pour la suite, P_x , R_x , C_x , V_x et Z_x représentent la valeur du profit, du revenu, des conversions, des visites et du coût de la meilleure solution trouvée, respectivement, où $x \in \{P, R, C, V\}$ où P, R, C, V sont utilisés pour maximisation de profit, maximisation de revenu, maximisation de conversion et maximisation de visite, respectivement. Par exemple, R_P est le revenu obtenu en maximisant les profits.

5.2.1 Comparaison des résultats avec différentes fonctions-objectifs

Toute campagne de marketing a un objectif. Il peut s'agir de se faire connaître, de promouvoir un produit ou service ou autres. Pour chaque objectif, l'annonceur choisira une ou plusieurs mesures de performance à observer pour déterminer s'il se rapproche de son objectif ou non. Toutefois, il arrive qu'il ne sache lequel choisir surtout quand son objectif est de tout simplement épuiser son budget tout en aillant un bon rendement. Cette problématique nous a motivé à analyser l'impact du choix de l'objectif sur les autres mesures de performance. Le but poursuivi ici est de trouver, si elle existe, la meilleure mesure parmi celles observées dans cette thèse qui peut être considérée comme un compromis c'est-à-dire dont l'optimisation n'a pas un lourd impact négatif sur les autres mesures (i.e. une détérioration de plus de 10% par rapport à leurs meilleures valeurs). Pour cela, la fonction-objectif du modèle initial a été remplacée par chacune de celles présentées dans la section 3.4.

Première vague : Comparaison entre Tabou₂ et Glouton

Dans une première vague, les tests ont été effectués sur les graphes G_1 à G_6 en utilisant le Tabou₂ et le glouton afin de comparer leurs performances. δ a été fixé à 0.05 pour cette vague. Les résultats sont présentés dans le tableau 5.5. Dans ce tableau, la variation de la valeur de l'objectif observé, noté ΔX , est égal à $X_{Gr} - X_{T2}$ où $X \in \{P, R, C, V\}$ et $T2$ représente le Tabou₂ et Gr le glouton.

Le tabou est généralement meilleur que le glouton. Toutefois, bien que la différence ici soit

Tableau 5.5 Changement d'objectif : Comparaison entre Tabou₂ et Glouton

Graphe	Objectif	Tabou ₂	Glouton	ΔX
G_1	Profits	-1137.56	-1137.56	0.00
G_1	Revenus	481.45	481.04	-0.41
G_1	Conversions	8.58	8.58	-0.01
G_1	Visites	1405.01	1390.57	-14.44
G_2	Profits	916.68	916.68	0.00
G_2	Revenus	1304.68	1299.06	-5.62
G_2	Conversions	23.54	23.46	-0.08
G_2	Visites	723.83	714.88	-8.95
G_3	Profits	-370.29	-370.29	0.00
G_3	Revenus	93.63	91.81	-1.82
G_3	Conversions	1.57	1.57	0.00
G_3	Visites	287.19	268.01	-19.18
G_4	Profits	713.23	713.23	0.00
G_4	Revenus	3147.01	3139.03	-7.98
G_4	Conversions	58.29	58.21	-0.08
G_4	Visites	1325.53	1311.72	-13.81
G_5	Profits	305.84	305.86	0.01
G_5	Revenus	740.62	739.79	-0.83
G_5	Conversions	16.61	16.49	-0.13
G_5	Visites	658.63	650.13	-8.50
G_6	Profits	506.44	507.53	1.09
G_6	Revenus	950.65	975.56	24.91
G_6	Conversions	16.49	16.18	-0.31
G_6	Visites	265.42	255.35	-10.08

plus marquée que dans la première série de tests, elle reste en bas de 5% sauf dans le cas de la maximisation de visites avec G_3 (où elle est plutôt en dessous de 7%). Nous notons quelques instances où le glouton a de meilleurs résultats que le tabou. Il est tout à fait possible que la précision choisie du tabou soit au moins une des raisons qui limitent sa performance tel que c'était le cas à la page 49. Par conséquent, le fait qu'il ait occasionnellement de moins bons résultats ne retient pas notre attention davantage d'autant plus que la différence observée ici est faible et que le temps de résolution est en moyenne 4 fois plus élevé avec le tabou comparativement au glouton.

Cela nous conforte dans l'idée de poursuivre nos tests uniquement avec le glouton.

Deuxième vague : Glouton et nouvelles instances

Dans la deuxième vague, les tests ont été effectués sur les nouvelles instances. Les résultats sont présentés dans les tableaux 5.6-5.10.

Dans le tableau 5.6, nous indiquons pour chaque instance la valeur de chaque mesure de performance à la meilleure solution trouvée dans le cas de la maximisation de profit : P_P , R_P , C_P et V_P , et, dans la dernière colonne, le coût de cette solution, Z_P . Ces résultats sont notre base de comparaison pour la suite. Ces résultats ont été obtenus avec le glouton combiné à la méthode de Gauss-Seidel optimisé. Ces algorithmes étant des heuristiques, il n'y a aucune garantie que les résultats qu'ils produisent soient optimaux. Néanmoins, comme nous le noterons, les écarts observés sont si faibles qu'il est raisonnable de croire que les observations faites dans cette section resteront valides avec les solutions optimales.

Tableau 5.6 Valeur des différentes mesures de performance à la meilleure solution trouvée pour la maximisation de profit

Graphe	P_P	R_P	C_P	V_P	Z_P
G_{13}	-23.57	1302.54	24.14	898.90	1326.11
G_{14}	4426.93	5364.00	96.45	1138.66	937.06
G_{15}	1176.38	1864.48	30.95	1428.54	688.10
G_{16}	-1086.58	453.97	8.14	3648.51	1540.56
G_{17}	346.07	2173.70	33.03	2143.54	1827.62
G_{18}	-18.30	5059.15	96.96	5868.71	5077.44
G_{19}	12299.53	17275.97	309.58	2774.00	4976.44
G_{20}	5691.64	6225.06	114.46	3562.91	533.42
G_{21}	10512.41	23782.82	426.54	4526.07	13270.41
G_{22}	25999.52	34828.61	603.09	6067.84	8829.10

Les tableaux 5.7-5.9 montrent la différence entre les résultats obtenus avec la maximisation de profit et ceux obtenus avec les autres fonctions-objectifs. Chaque colonne indique pour

chaque instance la variation de valeur d'une mesure de performance à la meilleure solution trouvée comparativement à celle trouvée dans le cas de la maximisation de profit notée $\Delta X_{x,P} = X_x - X_P$ où $X \in \{P, R, C, V, Z\}$ et $x \in \{P, R, C, V\}$. Autrement dit, elles montrent comment la valeur d'une mesure obtenue en maximisant le profit se compare à celle obtenue en maximisant cette mesure.

Dans le tableau 5.7, en regardant les valeurs de $\Delta P_{R,P}$, on observe que les profits sont grandement réduits lorsque l'objectif est de maximiser les revenus. En effet, en observant $\Delta Z_{R,P}$, on constate qu'il faut parfois dépenser en plus jusqu'à 30 fois (voir G_{15}) le montant de revenu supplémentaire. On pourrait attribuer cela au fait que maximiser les revenus nécessite de maximiser le nombre de visites aux pages menant aux conversions (à titre de rappel, seules les conversions génèrent des revenus et toute conversion est précédée par une page du site du client - voir section 3.2.2). Cela aura pour effet de mettre certains mots-clés à de meilleures positions ce qui sera plus coûteux alors que la maximisation des profits cherche à trouver le meilleur compromis entre coûts et revenus.

En observant $\Delta C_{R,P}$ et $\Delta V_{R,P}$, on remarque que la maximisation des revenus a généralement un meilleur impact sur le nombre de conversions et le nombre de visites que la maximisation des profits. Pour les quelques cas où c'est le contraire, la différence représente toujours moins de 5% de la meilleure valeur. Cela n'est pas très surprenant car revenus et conversions sont intimement liés : ce sont les conversions qui génèrent les revenus. De plus, pour parvenir aux conversions, et ainsi générer du revenu, les utilisateurs doivent généralement visiter le site donc il y a également un lien, bien qu'un peu moins direct, entre revenus et visites.

La surprise vient quand, en observant $\Delta R_{R,P}$, on constate que, même si on maximise les revenus, il arrive que la meilleure valeur trouvée soit moins bonne que celle trouvée lors de la maximisation des profits (voir G_{18} et G_{22}). Toutefois, ces différences représentent moins de 5% de la meilleure des deux valeurs. Cette observation n'est pas une exception. Comme nous l'observerons un peu plus tard, il arrive souvent qu'optimiser une mesure ne conduit pas nécessairement à la meilleure valeur observée pour cette mesure. Il est évident que ces anomalies sont tout simplement dues au fait que les solutions trouvées ne sont pas optimales. Toutefois, on peut se demander pourquoi on en observe autant. Afin de tenter de répondre à cette question, des tests supplémentaires ont été réalisés et sont présentés à la sous-section 5.2.1.

Tableau 5.7 Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de revenu comparativement à la maximisation de profit

Graphe	$\Delta P_{R,P}$	$\Delta R_{R,P}$	$\Delta C_{R,P}$	$\Delta V_{R,P}$	$\Delta Z_{R,P}$
G_{13}	-318.84	46.83	1.05	48.32	365.67
G_{14}	-238.92	13.46	0.30	26.06	252.38
G_{15}	-591.48	20.03	0.81	100.93	611.51
G_{16}	-691.50	28.30	0.43	3.30	719.80
G_{17}	-921.77	80.38	1.11	61.85	1002.14
G_{18}	-1090.63	-12.28	-0.90	-53.73	1078.35
G_{19}	-1019.69	203.14	2.90	42.59	1222.83
G_{20}	-1112.94	153.66	1.99	62.12	1266.60
G_{21}	-1202.25	408.91	6.27	134.15	1611.16
G_{22}	-3059.67	-1579.33	-23.36	-259.01	1480.34

Dans le tableau 5.8, en regardant les valeurs de $\Delta C_{C,P}$, on observe que maximiser les conversions est généralement moins efficace que de maximiser les profits bien que la différence soit inférieure à 5% de la meilleure des deux valeurs. De plus, en observant $\Delta Z_{C,P}$, on constate qu'il en coûte jusqu'à trois fois plus cher (voir G_{20}) pour avoir une moins bonne solution. Cette apparente anomalie est analysée avec les autres dans la sous-section suivante.

En observant $\Delta V_{C,P}$, on constate que la maximisation des conversions a généralement un meilleur impact sur le nombre de visites que la maximisation des profits. Pour les cas où c'est le contraire (voir G_{14} , G_{19} et G_{22}) ces différences représentent moins de 2% de la meilleure des deux valeurs. L'impact positif pourrait être expliqué par le lien qu'il y a entre les deux mesures. En effet, tel que noté précédemment, pour augmenter le nombre de conversions, il faut augmenter le nombre de visites aux pages du sites du client qui mènent à des conversions.

En observant $\Delta P_{C,P}$ et $\Delta R_{C,P}$, on remarque que la maximisation des conversions a généralement un pire impact sur les profits et les revenus que la maximisation des profits. La réduction de profits $\Delta P_{C,P}$ représente jusqu'à 63 fois la valeur obtenue avec la maximisation de profit. L'impact négatif sur les profits peut s'expliquer par le fait qu'il faille générer plus de visites pour générer plus de conversions et, dans notre système, le meilleur moyen de le faire est de placer des annonces à des positions plus élevées ce qui génère plus de coûts. Toutefois, comme la valeur des conversions n'est pas prise en compte ici, les revenus générés par ces conversions supplémentaires ne suivent certainement pas les coûts. Cela explique également pourquoi, bien qu'il y ait un lien positif entre conversions et revenus, la maximisation des conversions puisse avoir un impact négatif sur les revenus.

Tableau 5.8 Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de conversion comparativement à la maximisation de profit

Graphe	$\Delta P_{C,P}$	$\Delta R_{C,P}$	$\Delta C_{C,P}$	$\Delta V_{C,P}$	$\Delta Z_{C,P}$
G_{13}	-343.41	22.85	0.69	29.81	366.25
G_{14}	-411.50	-153.58	-1.73	-9.94	257.93
G_{15}	-695.19	-63.09	-0.38	89.99	632.10
G_{16}	-692.55	27.25	0.43	0.18	719.81
G_{17}	-1117.29	-111.82	-1.19	22.62	1005.47
G_{18}	-1155.84	-78.54	-0.82	127.15	1077.30
G_{19}	-1657.67	-393.64	-3.62	-47.66	1264.03
G_{20}	-1572.12	-383.55	-5.00	131.86	1188.57
G_{21}	-1936.51	-521.57	-5.26	102.80	1414.94
G_{22}	-1892.25	-617.52	-5.54	-54.80	1274.73

Dans le tableau 5.9, il est flagrant que le nombre de visites est la seule mesure qui est avantagée par la maximisation de visites avec une augmentation atteignant 10% (voir G_{17}). Toutefois, chaque visite supplémentaire coûte jusqu'à 5\$ l'unité (voir G_{19}).

En observant $\Delta C_{V,P}$, $\Delta P_{V,P}$ et $\Delta R_{V,P}$, on remarque que les dégradations sont nettement pires qu'avec n'importe laquelle des autres fonctions-objectifs. La maximisation des visites est donc la pire des quatre fonctions-objectifs en termes d'impact sur les autres mesures de performance. Ces dégradations peuvent s'expliquer par le fait que la maximisation de visites se résume à placer autant d'annonces que possible à de meilleures positions en privilégiant bien sûr celles qui peuvent en générer le plus. Ce qui fait grimper les coûts, ne tient compte ni du taux de conversions ni de la valeur des conversions et donc pas du revenu.

Tableau 5.9 Variation de valeur de chaque mesure de performance à la meilleure solution trouvée pour la maximisation de visite comparativement à la maximisation de profit

Graphe	$\Delta P_{V,P}$	$\Delta R_{V,P}$	$\Delta C_{V,P}$	$\Delta V_{V,P}$	$\Delta Z_{V,P}$
G_{13}	-377.27	-106.19	-1.91	69.19	271.08
G_{14}	-850.34	-616.04	-9.53	61.12	234.30
G_{15}	-923.03	-304.05	-4.26	150.69	618.98
G_{16}	-742.68	-24.37	-0.34	204.25	718.30
G_{17}	-1285.38	-313.47	-4.54	216.84	971.91
G_{18}	-1180.56	-356.37	-6.11	238.90	824.19
G_{19}	-2549.50	-1286.83	-21.32	236.45	1262.68
G_{20}	-1861.28	-814.28	-12.99	284.64	1047.01
G_{21}	-2541.26	-1307.21	-18.92	263.18	1234.05
G_{22}	-3107.53	-1900.89	-28.03	246.25	1206.64

Dans les tableaux 5.7-5.9, la comparaison était faite entre les résultats obtenus avec les nouvelles fonctions-objectifs et ceux obtenus avec la maximisation de profit. Dans le tableau 5.10, la comparaison est faite entre les résultats obtenus avec une des nouvelles fonctions-objectifs et ceux obtenus avec une autre. On y observe d'autres anomalies notamment avec $\Delta C_{R,C}$: on constate qu'il est généralement préférable d'utiliser la maximisation de revenus plutôt que la maximisation de conversions pour maximiser les conversions. Toutefois, la différence est très faible (toujours moins de 5%). Une autre exception est observée avec $\Delta R_{C,R}$ dans le cas de G_{22} mais la différence représente moins de 3% du revenu initial. Outre ces exceptions et celles observées auparavant, il est préférable et parfois même fortement recommandé d'utiliser la fonction-objectif qui maximise une mesure pour obtenir la meilleure valeur possible de cette mesure.

Tableau 5.10 Autres variations de valeur des mesures de performance à la meilleure solution trouvée

Graphe	$\Delta C_{R,C}$	$\Delta C_{V,C}$	$\Delta R_{C,R}$	$\Delta R_{V,R}$	$\Delta V_{C,V}$	$\Delta V_{R,V}$
G_{13}	0.36	-2.60	-23.98	-153.02	-39.39	-20.87
G_{14}	2.03	-7.80	-167.04	-629.50	-71.07	-35.07
G_{15}	1.19	-3.88	-83.12	-324.08	-60.70	-49.76
G_{16}	0.01	-0.77	-1.05	-52.67	-204.07	-200.96
G_{17}	2.30	-3.35	-192.20	-393.85	-194.22	-154.99
G_{18}	-0.07	-5.29	-66.26	-344.09	-111.74	-292.62
G_{19}	6.52	-17.70	-596.78	-1489.97	-284.11	-193.85
G_{20}	6.99	-7.99	-537.21	-967.94	-152.78	-222.52
G_{21}	11.53	-13.66	-930.48	-1716.12	-160.38	-129.02
G_{22}	-17.82	-22.49	961.81	-321.56	-301.04	-505.26

De plus, les résultats détaillés dans ces tableaux montrent que le profit et les coûts sont très sensibles au changement de fonction-objectif et que le revenu et les conversions sont impactées de la même façon. Maximiser les visites a l'impact le plus négatif sur toutes les autres mesures de performance. Dans ce modèle, maximiser les visites veut dire qu'on mise sur autant de mots-clés que possible pour générer autant de trafic que possible sur les pages du site de l'annonceur en privilégiant ceux avec le plus grand nombre de clics. Maximiser les conversions revient à miser autant que possible sur les mots-clés qui mènent aux pages du site de l'annonceur pour lesquelles le produit du taux de conversion et du nombre de clics est le plus grand. Maximiser le revenu c'est miser autant que possible sur les mots-clés qui mènent aux pages du site de l'annonceur pour lesquelles le produit du taux de conversion et de la valeur de la conversion est le plus grand. Maximiser le profit, pour sa part, consiste à trouver le meilleur équilibre entre les coûts et le revenu. Tout ceci pris en considération, il

est évident que la maximisation de profit est le meilleur compromis parmi ce qui est observé ici.

Analyses des anomalies

Plusieurs exceptions et irrégularités ont été observées dans les résultats des tests présentés précédemment. Leur existence a été attribuée au fait que les solutions ne sont pas optimales mais cela n'efface pas toutes les questions qu'elles soulèvent. Raison pour laquelle des tests complémentaires ont été effectués.

Pour confirmer qu'elles sont dues à l'existence d'une multitude d'extrema locaux, différents points de départ sont testés sur quatre des instances. À titre de comparaison, des irrégularités similaires ont été observées avec le tabou à quelques reprises : avec G_5 , en maximisant les revenus, on obtient un revenu de 740.62 mais en maximisant les conversions, le revenu est à 741.96 soit un écart de 0.18% ; avec G_6 , en maximisant les revenus, on obtient un revenu de 950.65 mais en maximisant le profit, le revenu était à 1002.10 soit un écart de 5% ; toujours avec G_6 , en maximisant les conversions, on obtient un nombre de conversions égal à 16.49 mais en maximisant le profit, le nombre de conversions est égal à 17.04 soit un écart de 3%.

Le but spécifique du test suivant est de déterminer si l'on obtient souvent différents résultats (rappel : l'algorithme glouton s'arrête lorsqu'il ne trouve plus de piste d'amélioration de la solution courante). Pour chaque graphe, dix scénarios sont testés. Dans le premier, tous les mots-clés commencent à la position 1. Il s'agit d'une solution extrême qui semble idéale : avec tous les mots-clés à la meilleure position et en ignorant la contrainte de budget, on tend à penser qu'on maximise le nombre de visites sur le site du client, le nombre de conversions ainsi que les revenus ce qui serait sans doute vrai si le flot à la source n'était pas fixe. Dans notre contexte, il est possible que dégrader la position de certains mots-clés contribue plus à améliorer au moins une de ces mesures que de le maintenir à sa position ou de l'améliorer. Dans l'exemple à la section 3.3, dans la situation initiale, $n_{w1} + n_{w2} + n_{w3} = 11$. Après avoir mis le mot-clé $k2$ à la position \bar{p} , la valeur de cette somme est de $\frac{6507}{507} = 12 + \frac{423}{507}$.

Dans le deuxième, ils commencent tous à la position $\bar{p} = 11$. Cette solution serait l'autre extrême : toutes les annonces textuelles sont désactivées, il ne reste que les bannières. On s'attend à être au plus bas du nombre de visites au site, du nombre de conversions, des coûts et des revenus. Tel qu'observé à l'exemple à la section 3.3, dans notre contexte, en mettant $k2$ à la position \bar{p} le nombre de visites au site n'est pas le seul à avoir augmenté, le nombre de visites à la bannière augmente aussi. Avec un coût suffisamment élevé (ce qui est généralement le cas dans l'industrie), cette augmentation de flot à la bannière entraînerait

une hausse des coûts totaux.

Pour les huit autres solutions de départ, les mots-clés commencent chacun à une position choisie aléatoirement dans $\mathcal{P} \setminus \{\bar{p}\}$. Pour chaque scénario, la différence ΔX entre la valeur de la mesure X^* à la meilleure solution trouvée et X_X est calculée. L'écart-type σ des valeurs observées est également calculé. Les résultats sont présentés dans le tableau 5.11.

Tableau 5.11 Différence entre la meilleure valeur de chaque mesure obtenue avec différents scénarios et celle obtenue avec le scénario initial

Graphe	Profit			Revenu			Conversions			Visites		
	Min	Max	σ	Min	Max	σ	Min	Max	σ	Min	Max	σ
G_{16}	0.00	0.00	0.00	-17.29	0.00	6.27	-0.41	0.00	0.13	0.00	51.16	20.22
G_{17}	0.00	0.00	0.00	-132.70	132.56	91.90	-1.31	0.88	0.79	-21.39	94.65	39.02
G_{19}	0.00	0.00	0.00	-812.63	95.98	392.45	-12.50	0.00	4.02	-42.94	143.63	61.83
G_{22}	-0.01	0.34	0.11	0.00	1906.82	511.53	-9.50	0.00	3.14	0.00	215.72	75.53

Ces résultats montrent que, dans le cas de la maximisation de profits, il y a très peu de variabilité ce qui laisse croire que, plus souvent qu'autrement, l'heuristique mène à l'optimum. La seule exception ici est avec le graphe G_{22} . On note d'ailleurs que d'autres exceptions avaient été observées avec ce même graphe (voir les tableaux 5.7, 5.8 et 5.10). Le graphe G_{18} est presque aussi gros avec presque autant d'arcs, de pertes et de visites et se distingue pour les mêmes exceptions. Il est possible qu'un de ces facteurs ou qu'une combinaison de ces facteurs aient un impact sur l'efficacité de l'algorithme.

Pour les autres fonctions-objectifs, la valeur de la meilleure solution trouvée peut varier selon la solution de départ mais cette variation est toujours strictement en-dessous de 6% de la valeur initiale. Ces observations permettent de ne pas éliminer la possibilité de présence de plusieurs optima locaux dans lesquels le glouton se retrouverait pris.

Une autre façon de tester cette hypothèse est de déterminer si l'ordre dans lequel chaque mot-clé est évalué a un impact sur la solution trouvée par l'algorithme. Pour cela, des tests ont été effectués avec un ordre différent : dans la version originale de l'algorithme, l'ordre est basé sur l'amélioration obtenue en changeant la position d'une unité ; alors que le nouvel ordre est simplement l'ordre lexicographique des identifiants des nœuds représentant les mots-clés dans le graphe. La différence entre les deux est présentée dans le tableau 5.12.

Tableau 5.12 Différence entre la meilleure valeur de chaque mesure obtenue avec différents scénarios et celle obtenue avec le scénario initial en utilisant l'ordre lexicographique

	Profit			Revenu			Conversions			Visites		
Graphe	Min	Max	σ	Min	Max	σ	Min	Max	σ	Min	Max	σ
G_{16}	0.00	0.00	0.00	-21.14	-10.58	3.01	-0.43	-0.21	0.07	-65.79	27.91	25.60
G_{17}	0.00	0.00	0.00	-127.99	100.73	75.15	0.38	2.42	0.65	-61.66	37.20	30.15
G_{19}	0.00	0.00	0.00	-739.54	79.57	297.19	-1.93	4.77	2.09	-128.34	42.99	44.99
G_{22}	0.25	0.59	0.14	1363.91	2045.19	236.85	-16.00	3.69	8.07	-57.43	79.29	49.07

Une fois de plus le graphe G_{22} se distingue avec une petite variance des résultats avec la maximisation des profits. L'hypothèse posée précédemment reste valide.

Il est intéressant de noter que les différences présentées au tableau 5.12 ne sont pas du tout les mêmes que celles présentées dans le tableau 5.11 ce qui renforce l'hypothèse selon laquelle il y a une multitude d'optima locaux.

5.2.2 Comparaison des résultats avec différents budgets

Dans cette partie de notre deuxième série de tests, une analyse de sensibilité sur le budget est faite. Dans ces tests, les meilleures solutions trouvées avec 80%, 90%, 100%, 110% et 120% du budget initial sont comparées. Les résultats sont compilés dans les tableaux 5.13-5.17.

Dans ce qui suit, nous notons la valeur du profit, du revenu, des conversions, des visites et du coût à la meilleure solution trouvée avec $y\%$ du budget, où $y \in \{80, 90, 110, 120\}$, avec P_x^y , R_x^y , C_x^y , V_x^y et Z_x^y , respectivement. Dans ces tableaux, $\Delta X_x^y = X_x^y - X_x$ où $x, X \in \{P, R, C, V, Z\}$ et $y \in \{80, 90, 11, 120\}$.

Dans le tableau 5.13, nous notons P_P , R_R , C_C et V_V la valeur de chaque mesure à la meilleure solution trouvée avec 100% du budget pour chaque instance, et $\frac{Z_x}{B}$, le pourcentage du budget dépensé pour cette solution. C'est la base de comparaison pour ce qui suit.

Dans les tableaux 5.14-5.17, nous indiquons ΔX_X^y et ΔZ_X^y , où $X \in \{P, R, C, V\}$ et $y \in \{80, 90, 110, 120\}$.

Tableau 5.13 Valeur à la meilleure solution trouvée pour chaque fonction-objectif et pourcentage du budget dépensé à chaque solution

Graphe	P_P	$\frac{Z_P}{B}$	R_R	$\frac{Z_R}{B}$	C_C	$\frac{Z_C}{B}$	V_V	$\frac{Z_V}{B}$
G_{13}	-23.57	78.34%	1349.37	99.94%	24.83	99.98%	968.10	94.36%
G_{14}	4426.93	77.74%	5377.46	98.67%	94.72	99.13%	1199.78	97.17%
G_{15}	1176.38	50.41%	1884.51	95.21%	30.57	96.72%	1579.23	95.76%
G_{16}	-1086.58	68.16%	482.27	100.00%	8.57	100.00%	3852.76	99.93%
G_{17}	346.07	64.49%	2254.07	99.84%	31.85	99.96%	2360.38	98.78%
G_{18}	-18.30	82.48%	5046.87	100.00%	96.14	99.98%	6107.60	95.87%
G_{19}	12299.53	79.73%	17479.11	99.33%	305.96	99.99%	3010.45	99.96%
G_{20}	5691.64	29.47%	6378.72	99.45%	109.46	95.14%	3847.55	87.32%
G_{21}	10512.41	89.17%	24191.73	99.99%	421.28	98.67%	4789.25	97.46%
G_{22}	25999.52	85.61%	33249.29	99.96%	597.55	97.97%	6314.09	97.31%

Tableau 5.14 Comparaison du profit dans le cas d'une maximisation de profit avec différents budgets

Graphe	ΔP_P^{80}	ΔP_P^{90}	ΔP_P^{110}	ΔP_P^{120}	$\frac{\Delta Z_P^{80}}{B}$	$\frac{\Delta Z_P^{90}}{B}$	$\frac{\Delta Z_P^{110}}{B}$	$\frac{\Delta Z_P^{120}}{B}$
G_{13}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{14}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{15}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{16}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{17}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{18}	-201.20	0.00	0.00	0.00	-0.69%	0.00%	0.00%	0.00%
G_{19}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{20}	0.00	0.00	0.00	0.00	0.00%	0.00%	0.00%	0.00%
G_{21}	-1014.60	0.00	0.00	0.00	-2.21%	0.00%	0.00%	0.00%
G_{22}	-1455.70	0.00	0.00	0.00	-4.26%	0.00%	0.00%	0.00%

Comme on pourrait s'y attendre, le tableau 5.14 montre qu'il n'y a généralement pas d'impact sur les profits car dans la majorité des cas, même pour la meilleure solution trouvée, les coûts représentent moins de 80% du budget. Pour les trois cas où les coûts étaient au-dessus de 80%, la coupure de 20% du budget a des cas où la meilleure solution était de garder toutes les annonces en position $\bar{p} = 11$ pour les graphes G_{18} , G_{21} et G_{22} .

Le tableau 5.15 montre que la variation des coûts suit généralement celle du budget. Toutefois, contrairement à la variation des coûts, la variation des revenus ne suit pas celle du budget de façon linéaire ni quasi-linéaire. Avec une augmentation de 10% du budget, la plus grande variation absolue des revenus (208.08) représente moins de 1% du meilleur revenu obtenu en maximisant les revenus pour le graphe correspondant (G_{22}); et avec une augmentation de 20% du budget, la plus grande variation de revenu (1966.71) représente moins de 6% du meilleur revenu pour le graphe correspondant (G_{22}). Le revenu est toutefois généralement plus sensible à une baisse de budget avec des variations. En effet, lorsque le budget diminue, les plus hautes variations observées (1595.34 et 1751.76 respectivement) représentent moins de 7% et 19% pour le graphe correspondant (G_{21}) respectivement.

On note également quelques cas où la diminution du budget entraîne une très légère augmentation de revenu (G_{14} , G_{15} et G_{17}) mais aussi le contraire (G_{21}). On peut supposer que la nouvelle limite de budget a affecté l'amplitude d'un changement de position, ce qui aurait modifié les changements subséquents menant éventuellement à de meilleures solutions.

Tableau 5.15 Comparaison du revenu dans le cas d'une maximisation de revenu avec différents budgets

Graphe	ΔR_R^{80}	ΔR_R^{90}	ΔR_R^{110}	ΔR_R^{120}	$\frac{\Delta Z_R^{80}}{B}$	$\frac{\Delta Z_R^{90}}{B}$	$\frac{\Delta Z_R^{110}}{B}$	$\frac{\Delta Z_R^{120}}{B}$
G_{13}	-128.42	-18.41	8.46	10.36	-19.95%	-10.11%	9.93%	17.93%
G_{14}	-369.48	9.49	27.22	76.86	-18.80%	-10.23%	7.04%	20.24%
G_{15}	-202.85	39.28	43.20	94.35	-15.42%	-8.71%	14.48%	20.47%
G_{16}	-5.86	-3.06	0.82	2.26	-20.00%	-10.00%	9.96%	15.54%
G_{17}	-406.11	5.30	25.10	26.18	-19.94%	-12.36%	9.94%	19.91%
G_{18}	-231.52	-109.47	150.43	122.85	-18.21%	-10.08%	9.98%	19.61%
G_{19}	-1224.71	-1164.59	3.24	24.11	-19.33%	-9.33%	5.11%	19.02%
G_{20}	-775.86	-688.06	12.21	11.02	-22.65%	-9.51%	7.15%	10.49%
G_{21}	-1751.76	-1595.34	-66.87	27.89	-13.03%	-10.00%	9.22%	18.47%
G_{22}	-315.78	-100.96	208.08	1966.71	-18.61%	-10.04%	9.91%	18.64%

Dans le tableau 5.16, on observe que les conversions sont moins sensibles que les revenus aux variations de budget. Avec 10% et 20% de plus au budget, les plus grandes variations sont de 6.09 et 8.38 respectivement. Ce qui représente seulement 6% et 8% du meilleur nombre de conversions obtenu en maximisant les conversions du graphe correspondant (G_{21}),

respectivement. Les conversions sont encore moins sensibles. Lorsque le budget varie de -10% et -20% , les variations observées sont de -4.94 et -22.21 respectivement. Cela représente moins de 2% et 4% pour les graphes correspondant (G_{21} et G_{22}), respectivement.

On note également quelques irrégularités comme avec le revenu (G_{16} et G_{17}) qui pourraient se justifier de la même façon.

Tableau 5.16 Comparaison du nombre de conversions dans le cas d'une maximisation de conversion avec différents budgets

Graphe	ΔC_C^{80}	ΔC_C^{90}	ΔC_C^{110}	ΔC_C^{120}	$\frac{\Delta Z_C^{80}}{B}$	$\frac{\Delta Z_C^{90}}{B}$	$\frac{\Delta Z_C^{110}}{B}$	$\frac{\Delta Z_C^{120}}{B}$
G_{13}	-1.44	-0.18	0.27	1.01	-20.23%	-10.69%	4.85%	20.02%
G_{14}	-0.18	-0.03	0.46	0.33	-19.94%	-9.85%	10.36%	20.66%
G_{15}	-0.89	-0.66	0.36	0.05	-16.74%	-7.97%	13.14%	8.36%
G_{16}	-0.10	-0.05	0.00	0.03	-20.00%	-10.01%	9.99%	15.32%
G_{17}	-2.86	1.22	0.08	-0.07	-20.01%	-18.91%	9.60%	19.99%
G_{18}	-2.76	-1.44	1.39	1.44	-18.19%	-10.41%	10.01%	20.02%
G_{19}	-10.62	-1.03	1.08	3.86	-19.99%	-10.84%	10.00%	19.99%
G_{20}	-1.49	-0.89	1.05	1.26	-17.43%	-7.27%	14.17%	24.48%
G_{21}	-15.07	-4.94	6.09	8.38	-11.71%	-8.69%	10.85%	21.28%
G_{22}	-22.21	-1.94	0.28	2.82	-16.62%	-9.15%	10.80%	19.33%

Les trois variations les plus grandes observées dans le tableau 5.17 sont des cas où la diminution de budget à résulter en des cas où il n'était pas possible de faire mieux que de n'afficher aucune annonce. Ces observations mises à part, les visites sont sans aucun doute les moins sensibles aux variations de budget. Les valeurs des variations obtenues sont faibles (toutes inférieures à 10 alors que le plus petit meilleur nombre de visites est 968.10). Cela correspond à des variations de moins de 0.5% . Ce qui mène à croire que baisser le budget autant que possible est une façon d'améliorer le retour sur investissement lorsqu'un annonceur souhaite attirer du trafic sur son site. Il faut toutefois garder en tête que selon les analyses de la sous-section précédente, la maximisation des visites a le pire impact sur les autres mesures de performance.

Pour résumer, l'impact est plutôt limité : une augmentation de 20% du budget ne se traduit pas en une amélioration de 20% de l'indice de performance même si le coût suit le budget. Cela est sans doute dû au fait que dans notre système, les coûts ne changent pas de façon linéaire mais exponentielle.

Tableau 5.17 Comparaison du nombre de visites dans le cas d'une maximisation de visite avec différents budgets

Graphe	ΔV_V^{80}	ΔV_V^{90}	ΔV_V^{110}	ΔV_V^{120}	$\frac{\Delta Z_V^{80}}{B}$	$\frac{\Delta Z_V^{90}}{B}$	$\frac{\Delta Z_V^{110}}{B}$	$\frac{\Delta Z_V^{120}}{B}$
G_{13}	-15.56	-3.32	4.30	5.94	-14.48%	-6.88%	10.09%	15.50%
G_{14}	-3.62	24.63	0.84	16.06	-17.79%	-7.18%	10.45%	22.75%
G_{15}	-2.09	-1.53	9.87	2.79	-17.30%	-8.87%	13.34%	17.39%
G_{16}	-8.36	9.79	5.14	10.60	-20.06%	-9.95%	9.83%	20.04%
G_{17}	-5.51	-1.98	2.57	29.32	-19.04%	-8.80%	9.12%	21.01%
G_{18}	-318.27	-0.72	8.64	67.51	-14.08%	-6.26%	14.09%	24.07%
G_{19}	-51.87	-15.44	26.11	45.73	-19.98%	-11.38%	10.00%	19.85%
G_{20}	0.71	-3.12	1.61	2.20	-8.12%	-8.91%	8.41%	16.61%
G_{21}	-526.93	-20.83	16.27	29.86	-10.50%	-7.48%	9.61%	22.10%
G_{22}	-520.95	-4.98	53.27	82.96	-15.96%	-8.37%	12.58%	22.64%

5.2.3 Comparaison des résultats avec différents CPCs

Dans la première série de tests, pour estimer le *CPC* moyen d'un mot-clé (*avgCPC*) à une nouvelle position, nous avons utilisé une formule déterministe mais, en réalité, il y a une marge d'erreur associée à toute estimation. Aussi, pour tester la robustesse du modèle face aux variations de l'*avgCPC*, dans cette dernière partie de notre seconde série de tests, les *avgCPCs* des mots-clés ont été modifiés aléatoirement selon la formule suivante :

$$cpc'_k = cpc_k (1 + 0.01rand) \quad \forall k \in \mathcal{K} \quad (5.1)$$

Où *rand* est un nombre aléatoire dans $\{-2(11 - p_k^{init}), \dots, 2(11 - p_k^{init})\}$. Le nouvel *avgCPC* se retrouve donc dans une marge de $\pm 20\%$ pour tout mot-clé initialement à la position 1 ; et l'amplitude de cette marge est progressivement réduite pour atteindre $\pm 2\%$ pour les mots-clés initialement à la position 10 tel qu'illustré dans la figure 5.1.

Dans ce qui suit, \mathcal{K}_c représente l'ensemble des mots-clés dont l'*avgCPC* a effectivement changé (c'est-à-dire tout mot-clé k tel que $cpc_k - cpc'_k \neq 0$) et \mathcal{K}_p , l'ensemble de mots-clés dont la position à la meilleure solution trouvée a changé. Pour chaque instance, $86\% \leq \frac{|\mathcal{K}_c|}{|\mathcal{K}|} \leq 90\%$ avec une moyenne de $88\% \pm 1.3\%$.

Les résultats sont présentés dans le tableau 5.18. On y retrouve, pour chaque graphe, le nombre de mots-clés dont l'*avgCPC* a effectivement changé ($|\mathcal{K}_p|$), la plus petite variation (Δp), la plus grande ($\Delta \bar{p}$) et la variation moyenne de l'*avgCPC* ($\Delta \hat{p}$).

Le profit est le moins sensible aux variations parmi les quatre mesures optimisées : moins de 7% des mots-clés changent de position et, le cas échéant, le changement n'est que d'une unité.

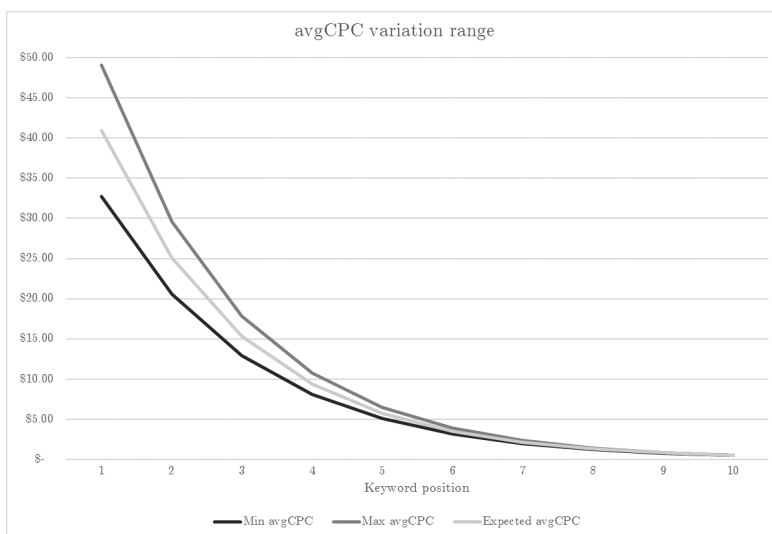


Figure 5.1 Amplitude de variation de l'*avgCPC* selon la position initiale d'un mot-clé : cas d'un mot-clé dont l'*avgCPC* initial est de 2\$ à la position 10

Pour les autres mesures, l'amplitude des changements est plus grande et le changement moyen généralement plus important. Pour la maximisation des revenus, jusqu'à 47% des mots-clés changent de position. Cette proportion est de 29% pour la maximisation des conversions et 72% pour la maximisation des visites. Il est évident que cette dernière est la plus sensible aux variations de *CPC*, ce qui peut sembler assez surprenant car elle était la moins sensible à la variation de budget. Toutefois, le nombre de visites est sans aucun doute la mesure étudiée ici qui est le lien le plus direct avec le *CPC*. En effet, plus le *CPC* est élevé plus il est coûteux d'améliorer une position et plus la dégradation d'une position libérera du budget, le contraire si le *CPC* baisse. Le nombre de visites est directement liée aux positions des annonces. Il est donc normal qu'une variation de *CPC* qui a un impact direct sur les positions des annonces causent une réorganisation massive de celles-ci.

Tableau 5.18 Sommaire des variations des positions des mots-clés

Fonction-objectif	Graphe	$ \mathcal{K}_p $	(Δp)	$(\Delta \bar{p})$	$(\Delta \hat{p})$
Profit	G_{13}	10	-1	1	0.00
	G_{14}	6	-1	1	0.00
	G_{15}	5	-1	1	-0.60
	G_{16}	4	-1	1	0.00
	G_{17}	22	-1	1	-0.18
	G_{18}	22	-1	1	-0.18
	G_{19}	32	-1	1	0.13
	G_{20}	16	-1	1	-0.25
	G_{21}	37	-1	1	-0.19
	G_{22}	27	-1	1	-0.11
Revenu	G_{13}	7	-4	4	0.14
	G_{14}	71	-10	10	-0.35
	G_{15}	43	-6	5	-0.05
	G_{16}	3	-3	-1	-2.00
	G_{17}	21	-3	3	-0.19
	G_{18}	21	-3	3	-0.19
	G_{19}	29	-5	7	0.28
	G_{20}	20	-3	1	-0.40
	G_{21}	197	-9	10	0.51
	G_{22}	2	1	1	1.00
Conversions	G_{13}	43	-10	8	-1.35
	G_{14}	5	-1	7	1.80
	G_{15}	23	-6	2	0.17
	G_{16}	4	1	5	2.75
	G_{17}	9	-2	4	0.78
	G_{18}	9	-2	4	0.78
	G_{19}	21	-6	4	-0.43
	G_{20}	57	-9	7	-0.04
	G_{21}	56	-9	10	-0.34
	G_{22}	38	-2	2	-0.11
Visites	G_{13}	13	-5	1	0.08
	G_{14}	13	-1	1	-0.08
	G_{15}	96	-9	1	-1.56
	G_{16}	217	-8	9	-1.82
	G_{17}	129	-10	10	-1.27
	G_{18}	129	-10	10	-1.27
	G_{19}	81	-8	9	0.25
	G_{20}	52	-3	1	-0.02
	G_{21}	78	-4	3	-0.17
	G_{22}	57	-1	1	0.02

Le tableau 5.19 montre un extrait des variations de la répartition des mots-clés par position. Cet extrait montre les cas avec les variations les plus extrêmes. Ce ne sont pas les cas avec les plus grands nombres de changement mais ceux avec les changements qui ont la plus grande amplitude. On note que les changements les plus importants affectent les positions extrêmes (les deux premières et les deux dernières). Comme on peut y voir, pour la maximisation de profit, après que les *avgCPCs* aient été changés, il n'y a qu'une légère redistribution des positions aucun changement majeur. Pour les autres fonctions-objectifs, par contre, on compte toujours au moins un cas avec des changements importants :

- Maximisation des conversions : G_{13} positions 1, 2, 10 et 11 ;
- Maximisation de revenu : G_{14} positions 1, 2 et 10 ;
- Maximisation des visites : G_{15} positions 1, 10 et 11.

Les tableaux B.1 et B.2 de l'annexe B contiennent les résultats pour tous les graphes. Dans la majorité des cas, la répartition reste essentiellement la même ($\pm 2\%$) mais, comme on peut le voir dans le tableau 5.18, il peut y avoir jusqu'à 217 mots-clés qui ont changé de positions et les variations peuvent être extrêmes (± 10 unités sauf dans le cas de la maximisation de profit).

Dans le tableau 5.20, sont présentées les variations de l'*avgCPC* pour chaque variation de position pour quelques graphes. Les valeurs en rouge sont les variations de l'*avgCPC* qui ont un signe différent de celui de la variation de position correspondante. Juste avec cet extrait, il peut être déduit qu'il n'y a aucune corrélation entre la variation de position Δp et la variation de l'*avgCPC* (une fois de plus, à l'exception de la maximisation de profit). En effet, il arrive souvent que des mots-clés devenus moins coûteux se retrouvent à de moins bonnes positions ou, à l'inverse, que des mots-clés plus coûteux se retrouvent à de meilleures positions. Ceci n'est pas une parfaite surprise car, au final, le seul lien entre les coûts et l'objectif c'est le budget. L'hypothèse peut être faite que les mots-clés moins coûteux ont permis de libérer suffisamment de budget pour que certains mots-clés plus performants bien que plus coûteux puissent profiter de positions plus avantageuses.

Les tableaux B.3 et B.4 de l'annexe B contiennent les résultats pour tous les graphes.

Tableau 5.19 Variation de la proportion de mots-clés à toutes les positions à la meilleure solution trouvée après la variation des *avgCPCs*

	Position	G_{13}	G_{14}	G_{15}
Profit	1	0%	-1%	0%
	2	1%	1%	0%
	3	-1%	0%	0%
	4	1%	1%	0%
	5	-2%	-1%	1%
	6	0%	1%	-1%
	7	1%	0%	0%
	8	0%	1%	0%
	9	-1%	-1%	0%
	10	1%	0%	0%
	11	0%	0%	0%
Revenu	1	0%	16%	0%
	2	0%	-15%	0%
	3	-1%	0%	-1%
	4	1%	-1%	0%
	5	0%	1%	1%
	6	1%	1%	0%
	7	1%	-1%	2%
	8	0%	-1%	0%
	9	-1%	-2%	-2%
	10	1%	4%	0%
	11	-1%	-1%	0%
Conversions	1	5%	-1%	-1%
	2	-4%	0%	0%
	3	1%	0%	1%
	4	1%	0%	0%
	5	1%	-1%	0%
	6	2%	1%	0%
	7	0%	-1%	-1%
	8	0%	1%	1%
	9	-2%	1%	1%
	10	3%	0%	0%
	11	-7%	0%	0%
Visites	1	-1%	0%	11%
	2	1%	0%	-1%
	3	0%	0%	-2%
	4	0%	-1%	0%
	5	0%	0%	-2%
	6	0%	1%	-1%
	7	-1%	-1%	-1%
	8	1%	0%	-1%
	9	1%	1%	-2%
	10	-1%	-1%	-13%
	11	0%	0%	11%

Tableau 5.20 Variation de l'*avgCPC* pour chaque variation de position

	Δp	Graphe 1	Graphe 2	Δp	Graphe 7
Profit	1	13.00%	8.00%	-1	-10.07%
	2	-	-	-2	-
	3	-	-	-3	-
	4	-	-	-4	-
	5	-	-	-5	-
	6	-	-	-6	-
	7	-	-	-7	-
	8	-	-	-8	-
	9	-	-	-9	-
	10	-	-	-10	-
Revenu	1	13.00%	0.00%	-1	-8.40%
	2	-	1.00%	-2	13.00%
	3	-	9.00%	-3	-
	4	6.00%	-3.00%	-4	-
	5	-	-2.00%	-5	-11.00%
	6	-	1.00%	-6	-
	7	-	-	-7	-
	8	-	13.00%	-8	-
	9	-	-3.00%	-9	-
	10	-	-3.67%	-10	-
Conversions	1	0.21%	13.67%	-1	-6.91%
	2	-	-	-2	9.00%
	3	2.00%	-	-3	-
	4	-	-	-4	-
	5	-3.00%	-	-5	-
	6	-	-	-6	15.00%
	7	-	3.00%	-7	-
	8	-1.00%	-	-8	-
	9	-	-	-9	-
	10	-	-	-10	-
Visites	1	11.11%	13.50%	-1	-6.50%
	2	-	-	-2	-5.00%
	3	-	-	-3	15.00%
	4	-	-	-4	-
	5	-	-	-5	0.67%
	6	-	-	-6	13.00%
	7	-	-	-7	11.00%
	8	-	-	-8	0.67%
	9	-	-	-9	-
	10	-	-	-10	-

CHAPITRE 6 CONCLUSION

6.1 Synthèse des travaux

La gestion d'une campagne marketing en ligne est un processus complexe et sensible à un grand nombre de variables. La difficulté est d'autant plus grande que les annonceurs n'ont généralement accès qu'à une quantité limitée d'informations concernant leurs campagnes. Notre objectif était de mettre à profit ces informations au maximum pour améliorer la performance des campagnes. Notamment en utilisant les historiques de navigation des utilisateurs. Aussi, dans cette thèse, nous avons présenté un modèle d'optimisation des campagnes de marketing interactifs. Le modèle est inspiré du problème de flot maximum (ou du problème de flot à coût minimum selon l'objectif utilisé) mais n'est pas linéaire. Il a donc fallu concevoir un algorithme de résolution. Toutefois, le flot dans notre problème dépend des positions des annonces. Ce qui soulève le défi de formuler les contraintes de conservation de flot cohérentes mais aussi, permet d'aborder le problème en deux étapes. La première étape consiste à choisir les positions des annonces et la seconde à calculer le flot résultant.

Pour la première étape, nous nous sommes basés sur des algorithmes existants pour développer nos algorithmes de résolution en commençant par le tabou. Bien que le temps de résolution fût trop important, cela nous a permis de réaliser qu'une descente était envisageable. Dans une seconde approche, nous avons développé un algorithme glouton. Cet algorithme consiste, à partir d'une solution donnée, à évaluer l'impact d'un changement d'une unité de la position de chaque annonce. Il retient le changement le plus profitable pour chacune des annonces, s'il y en a, sinon maintient l'annonce à la même position puis applique chaque changement l'un après l'autre en commençant par le plus profitable dans l'ensemble. L'algorithme répète le processus jusqu'à ce qu'il n'y ait plus d'amélioration possible. Cet algorithme s'est avéré beaucoup plus rapide que le tabou. De plus, durant les tests, nous avons constaté que partir d'une solution où aucune des annonces ne serait affichée permettait de gagner encore plus de temps. Ceci est certainement lié au fait que la meilleure solution était très souvent de laisser de nombreuses annonces non affichées.

Pour la deuxième étape, il suffit de résoudre un système d'équations. Initialement, nous laissons la tâche à CPLEX mais nous avons réalisé que les méthodes itératives étaient beaucoup plus rapides. Celle qui s'est particulièrement démarquée est sans aucun doute la forme optimisée de la méthode de Gauss-Seidel.

Dans la seconde partie de notre travail, nous avons effectué une analyse comparative de l'impact de divers changements sur les solutions. Premièrement, nous avons modifié la fonction-

objectif du modèle, appliqué l’algorithme glouton et comparé les solutions obtenues avec chacune des fonctions-objectifs. Nous avons constaté que les profits étaient très sensibles à ce changement et que la maximisation des visites avait le plus grand impact négatif sur les autres indicateurs de performance. En effet, les fonctions-objectifs considérées autres que la maximisation de profit ont tendance à placer les mots-clés à des positions plus élevées ; c’est particulièrement le cas de la maximisation des visites.

Nous avons également analysé l’impact d’une variation de budget. Les profits y sont peu sensibles contrairement aux autres mesures de performance. Toutefois, pour eux, la variation de valeur ne semble pas être directement proportionnelle à la variation de budget. Toutefois, nous avons également constaté que, contrairement à la maximisation des profits, notre algorithme glouton ne converge pas toujours vers la même solution.

Bref, nous avons développé un modèle permettant de répondre au besoin des annonceurs en exploitant les informations qui leur sont accessibles dans le but d’améliorer la performance de leurs campagnes. Lorsque plusieurs campagnes sont représentées, il optimise aussi la répartition du budget à travers les différentes campagnes. Le tout en optimisant un indice de performance. Il présente l’avantage d’être flexible.

6.2 Limitations de la solution proposée

Dans une campagne interactive, les positions des bannières sont également mises en enchère. Dans le cadre de ce travail, les positions des bannières ont été considérées comme constantes, l’impact des modifications sur les futures interactions n’ont pas été prises en compte et le modèle est déterministe alors que la plupart des paramètres varient dans le temps. De plus, l’algorithme glouton présenté dans cette thèse produit des résultats non optimaux. Ce qui pourrait expliquer la plupart sinon toutes les anomalies observées dans les résultats des tests.

6.3 Améliorations futures

Plusieurs pistes d’améliorations ressortent de certaines limitations. La première est de tenir compte de l’impact du changement de position des bannières. Une approche serait de trouver des fonctions génériques pour les bannières, si elles existent.

Une autre amélioration possible serait de compléter le modèle en y ajoutant l’impact des changements de position sur les futures interactions des utilisateurs. Ce modèle pourrait être intégré au modèle d’optimisation et l’objectif serait alors d’optimiser la performance espérée d’une campagne sur un horizon donné.

Il y a certainement place à l'amélioration de la méthode de résolution telle que l'usage d'autres algorithmes de recherche locale ou d'algorithmes de recherche à voisinages variables seuls ou en combinaison.

RÉFÉRENCES

- N. Archak, V. Mirrokni, et S. Muthukrishnan, “Budget optimization for online advertising campaigns with carryover effects”, dans *Sixth Ad Auctions Workshop*. Citeseer, 2010.
- K. Azeuli, M. Gamache, A. Hertz, et S. Paroz, “Optimizing keyword positions for search engine marketing”, Les Cahiers du GERAD, 2017.
- C. Borgs, J. Chayes, N. Immorlica, K. Jain, O. Etesami, et M. Mahdian, “Bid optimization in online advertisement auctions”, dans *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 531–540.
- D. Bosomworth. (2016) Kpis for measuring content marketing roi. En ligne : <http://www.smartinsights.com/content-management/content-marketing-strategy/measuring-content-marketing-kpis/>
- S. Brady. (2016) What percent of revenue do publicly traded companies spend on marketing and sales? En ligne : <https://vtldesign.com/inbound-marketing/content-marketing-strategy/percent-of-revenue-spent-on-marketing-sales/>
- N. Brooks, “The atlas rank report : How search engine rank impacts traffic”, *Insights, Atlas Institute Digital Marketing*, 2004.
- N. Brooks, “The atlas rank report—part ii : How search engine rank impacts conversion”, *Seattle : Atlas Institute*, 2004.
- B. DasGupta et S. Muthukrishnan, “Stochastic budget optimization in internet advertising”, *Algorithmica*, vol. 65, no. 3, pp. 634–661, 2013.
- P. S. Desai, W. Shin, et R. Staelin, “The company that you keep : when to buy a competitor’s keyword”, *Marketing Science*, vol. 33, no. 4, pp. 485–508, 2014.
- C. Edgecomb. The 10 marketing kpis you should be tracking. En ligne : <http://www.impactbnd.com/the-10-marketing-kpis-you-should-be-tracking>
- E. Even Dar, V. S. Mirrokni, S. Muthukrishnan, Y. Mansour, et U. Nadav, “Bid optimization for broad match ad auctions”, dans *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 231–240.

- J. Feldman, S. Muthukrishnan, M. Pal, et C. Stein, “Budget optimization in search-based advertising auctions”, dans *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 2007, pp. 40–49.
- J. Feng, H. K. Bhargava, et D. M. Pennock, “Implementing sponsored search in web search engines : Computational evaluation of alternative mechanisms”, *INFORMS Journal on Computing*, vol. 19, no. 1, pp. 137–148, 2007.
- A. Fortin, *Analyse numérique pour ingénieurs*. Presses inter Polytechnique, 2011.
- M. Gendreau, A. Hertz, et G. Laporte, “A tabu search heuristic for the vehicle routing problem”, *Management Science*, vol. 40, pp. 1276–1290, 1994.
- F. Glover, “Future paths for integer programming and links to artificial intelligence”, *Computers & Operations Research*, vol. 5, pp. 533–549, 1986.
- F. Glover, “Tabu search—part i”, *ORSA Journal on computing*, vol. 1, no. 3, pp. 190–206, 1989.
- F. W. Glover et M. Laguna, *Tabu Search*. Springer, 1997.
- T. Graepel, J. Q. Candela, T. Borchert, et R. Herbrich, “Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft’s bing search engine”, dans *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 13–20.
- C. Karande, A. Mehta, et R. Srikant, “Optimizing budget constrained spend in search advertising”, dans *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 697–706.
- B. Kitts et B. Leblanc, “Optimal bidding on keyword auctions”, *Electronic Markets*, vol. 14, no. 3, pp. 186–201, 2004.
- H.-F. Li, S.-Y. Lee *et al.*, “Mining top-k path traversal patterns over streaming web click-sequences”, *Journal of Information Science and Engineering*, vol. 25, no. 4, pp. 1121–1133, 2009.
- S. Lu, Y. Zhu, et A. Dukes, “Position auctions with budget constraints : Implications for advertisers and publishers”, *Marketing Science*, vol. 34, no. 6, pp. 897–905, 2015.

- H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin *et al.*, “Ad click prediction : a view from the trenches”, dans *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1222–1230.
- O. Nasraoui et R. Krishnapuram, “One step evolutionary mining of context sensitive associations and web navigation patterns”, dans *Proceedings of the 2002 SIAM International Conference on Data Mining*. SIAM, 2002, pp. 531–547.
- O. Nasraoui, C. Cardona, C. Rojas, et F. Gonzalez, “Mining evolving user profiles in noisy web clickstream data with a scalable immune system clustering algorithm”, dans *Proc. of WebKDD*, 2003, pp. 71–81.
- P. Quinn, “Modélisation et prédiction du comportement de mots-clés dans des campagnes publicitaires sur les moteurs de recherche”, Mémoire de maîtrise, École Polytechnique de Montréal, 2011.
- O. J. Rutz, R. E. Bucklin, et G. P. Sonnier, “A latent instrumental variables approach to modeling keyword conversion in paid search advertising”, *Journal of Marketing Research*, vol. 49, no. 3, pp. 306–319, 2012.
- B. Skiera et N. Abou Nabout, “Practice prize paper—prosad : a bidding decision support system for profit optimizing search engine advertising”, *Marketing Science*, vol. 32, no. 2, pp. 213–220, 2013.
- Y. Yang, J. Zhang, R. Qin, J. Li, F.-Y. Wang, et W. Qi, “A budget optimization framework for search advertisements across markets”, *Systems, Man and Cybernetics, Part A : Systems and Humans, IEEE Transactions on*, vol. 42, no. 5, pp. 1141–1151, 2012.
- S. Yao et C. F. Mela, “A dynamic model of sponsored search advertising”, *Marketing Science*, vol. 30, no. 3, pp. 447–468, 2011.
- W. Zhang, Y. Zhang, B. Gao, Y. Yu, X. Yuan, et T.-Y. Liu, “Joint optimization of bid and budget allocation in sponsored search”, dans *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1177–1185.
- Y. Zhou, D. Chakrabarty, et R. Lukose, “Budget constrained bidding in keyword auctions and online knapsack problems”, dans *Internet and Network Economics*. Springer, 2008, pp. 566–576.

ANNEXE A GLOSSAIRE

Afficheur (*publisher*) L’afficheur est toute personne ou tout organisme qui vend et gère un espace d’affichage d’une publicité.

Annonce organique (*organic ad*) Une annonce organique est le résultat d’une recherche textuelle extrait d’un site Web et non payé (à distinguer de l’annonce publicitaire). Ce type d’annonces n’existe que sur les moteurs de recherche.

Annonce publicitaire (*ad*) Une annonce publicitaire est une annonce pouvant avoir divers formats (voir Types de publicité à la section 1.1.1) créé par un annonceur dont il se sert pour attirer le public.

Annonceur (*advertiser*) L’annonceur est toute personne ou tout organisme qui lance et gère une campagne de publicité.

Bannière (*banner*) Une bannière est l’emplacement d’une publicité graphique.

Chemin de conversion (*conversion path*) Une chemin de conversion est une portion d’historique de navigation d’un utilisateur commençant par une interaction (par exemple, recherche dans un moteur de recherche, clic sur une bannière ou sur un lien reçu par courriel, etc.) et se terminant par une conversion ou par une autre interaction. Une conversion ne peut être qu’à la fin d’un chemin. Si l’historique de navigation de l’utilisateur continue, la suite sera représentée dans un autre chemin.

Conversion (*conversion*) Une conversion est une action que l’annonceur veut que l’utilisateur réalise après avoir été exposé à ses publicités. Il peut s’agir d’une inscription à une infolettre, du remplissage d’un formulaire, d’une demande d’information ou d’un achat. Dans l’historique de navigation, il s’agit de la page du site web de l’annonceur sur laquelle est affichée la confirmation que l’utilisateur a complété l’action prédéfinie. Un site peut avoir plus d’une page de confirmation (une par type d’actions par exemple).

Créatif (*creative*) Un créatif est une version d’une annonce publicitaire. Dans le cadre d’une campagne publicitaire, l’annonceur a la possibilité de séparer ses mots-clés en plusieurs

groupes (voir groupe d'annonces) et d'attribuer à chaque groupe un ou plusieurs créatifs en plus d'autres spécificités. Cela a pour avantage d'afficher du contenu plus personnalisé en fonction de l'utilisateur.

Enchère ou *CPC* max (*bid or max CPC*) Une enchère, plus communément appelée *CPC* max dans l'industrie de la publicité en ligne, est un montant maximal qu'un annonceur est prêt à payer pour que son annonce soit affichée.

Groupe d'annonces (*adgroup*) Un groupe d'annonces est un ensemble de mots-clés, créatifs et de caractéristiques servant à cibler un public dans le cadre d'une campagne publicitaire. Chaque campagne en comporte généralement au moins deux et les mots-clés de chaque groupe sont distincts (un mot-clé d'un groupe ne se retrouve pas dans l'autre) ainsi, si l'annonceur est sélectionné pour afficher son annonce, celle-ci proviendra de l'un des groupes.

Impression (*impression*) Une impression est un affichage d'une annonce publicitaire donc chaque fois qu'une annonce apparaît sur l'écran d'un utilisateur, cela compte comme une impression.

Indice de qualité (*quality score*) L'indice de qualité d'une annonce évalue la relation entre les mots-clés, les annonces et les pages d'atterrissage. L'indice de qualité est affecté aux mots-clés. Un indice de qualité élevé signifie que l'afficheur estime que l'annonce, le mot-clé et la page d'atterrissage sont adéquats et utiles aux personnes qui cliquent sur l'annonce. "Le calcul de l'indice de qualité varie légèrement d'un moteur de recherche à l'autre, mais il est fortement influencé par le taux de clic des annonces" (Quinn, 2011).

Modèle d'attribution des conversions Le modèle d'attribution des conversions est un modèle mathématique servant à attribuer un poids à chaque événement d'un chemin de conversion d'un utilisateur.

Mot-clé (*keyword*) Un mot-clé est un mot ou une expression choisi(e) par l'annonceur et qui, selon lui, a des chances d'être utilisé par des utilisateurs susceptibles d'être intéressés par son site.

Moteur de recherche (*search engine*) Un moteur de recherche est un "outil informatique qui [...] permet de repérer et d'accéder à des pages Web, images, vidéos et autres types de fichiers correspondant à des critères de recherche spécifiques. Les moteurs de recherche les

plus efficaces sont ceux qui réussissent à fournir, en peu de temps, les résultats associés à une requête donnée, classés en ordre décroissant de pertinence" (Quinn, 2011).

Marketing interactif Une campagne de marketing interactif est toute campagne de publicité qui note la réaction de la personne visée après son exposition à un ou plusieurs éléments de la campagne et se réajuste suite à cette réaction.

Navigateur (*web browser*) Un navigateur est un logiciel permettant à un utilisateur de parcourir des sites Web.

Paire mot-clé et annonce textuelle Un mot-clé peut être lié à plus d'une annonce textuelle et une annonce peut être liée à plus d'un mot-clé. C'est pourquoi dans cette thèse, nous ferons référence à la paire mot-clé - annonce plutôt que de les désigner séparément.

Référencement naturel (*web indexing*) Il s'agit d'une méthode d'indexation du contenu d'un site web. Les moteurs de recherche le font en utilisant des mots-clés ou des métadonnées ce qui leur permet d'identifier des contenus pertinents selon la requête d'un utilisateur.

Requête (*query*) Une requête est un ensemble de mots utilisés par un utilisateur sur un moteur de recherche pour décrire ce qu'il recherche.

Taux de clic (*click-through-rate or CTR*) Le taux de clic est le rapport du nombre de clics sur une annonce par le nombre d'impressions de cette même annonce.

Utilisateur (*user*) Un utilisateur est toute personne qui, durant sa navigation sur internet, a interagi avec au moins un élément de la campagne publicitaire de l'annonceur. Dans le cadre de ce travail, la seule interaction considérée est le clic.

ANNEXE B TABLEAUX COMPLETS

Tableau B.1 Proportion de mots-clés à chaque position à la meilleure solution trouvée avant et après la variation des *avgCPCs* - Graphe 1 à 5

		Graphe 1		Graphe 2		Graphe 3		Graphe 4		Graphe 5	
	Position	Avant	Après	Avant	Après	Avant	Après	Avant	Après	Avant	Après
Profit	1	2%	2%	5%	4%	2%	2%	0%	0%	2%	2%
	2	2%	3%	1%	1%	1%	1%	0%	0%	2%	2%
	3	2%	1%	3%	3%	1%	1%	0%	0%	2%	2%
	4	3%	4%	4%	5%	3%	2%	1%	0%	3%	3%
	5	4%	2%	3%	2%	3%	4%	1%	1%	2%	1%
	6	4%	4%	4%	5%	3%	2%	0%	0%	4%	3%
	7	2%	3%	2%	2%	4%	4%	1%	2%	3%	3%
	8	4%	4%	3%	4%	5%	5%	1%	1%	5%	4%
	9	3%	3%	5%	5%	4%	4%	1%	1%	5%	6%
	10	12%	13%	7%	7%	10%	10%	9%	9%	12%	11%
	11	62%	62%	63%	63%	64%	64%	85%	85%	62%	62%
Revenu	1	11%	11%	7%	23%	6%	6%	2%	2%	4%	4%
	2	16%	16%	16%	1%	11%	10%	0%	0%	2%	2%
	3	4%	3%	1%	1%	3%	2%	0%	1%	2%	1%
	4	1%	2%	3%	1%	2%	2%	0%	0%	3%	3%
	5	0%	0%	1%	1%	4%	5%	0%	0%	2%	2%
	6	1%	2%	1%	2%	5%	5%	0%	0%	2%	2%
	7	5%	5%	3%	3%	3%	5%	0%	0%	2%	3%
	8	4%	4%	4%	3%	4%	4%	0%	0%	4%	4%
	9	5%	4%	5%	3%	6%	4%	0%	0%	3%	3%
	10	11%	12%	11%	15%	19%	19%	0%	0%	28%	26%
	11	42%	41%	49%	47%	36%	36%	96%	95%	48%	48%
Conversions	1	1%	5%	1%	1%	5%	5%	1%	1%	1%	1%
	2	15%	11%	1%	1%	11%	11%	1%	1%	1%	0%
	3	1%	1%	0%	0%	1%	2%	0%	0%	0%	1%
	4	1%	3%	1%	1%	2%	2%	1%	0%	1%	1%
	5	3%	3%	2%	1%	2%	3%	0%	0%	0%	0%
	6	4%	6%	1%	2%	3%	3%	1%	0%	2%	2%
	7	3%	3%	2%	1%	5%	3%	0%	0%	1%	2%
	8	2%	2%	5%	5%	3%	3%	0%	0%	1%	1%
	9	8%	6%	1%	1%	3%	4%	0%	0%	1%	2%
	10	13%	17%	29%	29%	27%	27%	0%	0%	40%	40%
	11	49%	43%	58%	58%	38%	37%	96%	97%	51%	51%
Visites	1	5%	4%	3%	3%	3%	14%	11%	26%	4%	10%
	2	1%	2%	2%	2%	2%	1%	2%	1%	1%	2%
	3	4%	4%	3%	3%	4%	1%	2%	3%	4%	3%
	4	2%	2%	5%	5%	4%	5%	5%	6%	5%	4%
	5	9%	9%	6%	6%	6%	4%	5%	4%	7%	6%
	6	11%	11%	11%	13%	8%	7%	7%	8%	10%	8%
	7	9%	9%	15%	15%	9%	8%	7%	5%	11%	10%
	8	8%	9%	5%	5%	10%	9%	7%	7%	9%	8%
	9	7%	7%	5%	7%	10%	8%	10%	14%	8%	7%
	10	44%	43%	43%	41%	45%	31%	36%	16%	41%	34%
	11	0%	0%	0%	0%	0%	11%	8%	10%	0%	8%

Tableau B.2 Proportion de mots-clés à chaque position à la meilleure solution trouvée avant et après la variation des *avgCPCs* - Graphe 6 à 10

		Graphe 6		Graphe 7		Graphe 8		Graphe 9		Graphe 10	
	Position	Avant	Après	Avant	Après	Avant	Après	Avant	Après	Avant	Après
Profit	1	2%	2%	4%	4%	2%	2%	3%	3%	4%	4%
	2	2%	2%	4%	4%	2%	2%	1%	1%	3%	3%
	3	2%	2%	4%	3%	2%	2%	3%	3%	3%	3%
	4	3%	3%	3%	4%	1%	1%	4%	4%	3%	3%
	5	2%	1%	4%	4%	4%	3%	3%	3%	4%	4%
	6	4%	3%	3%	3%	3%	3%	6%	5%	5%	4%
	7	3%	3%	4%	4%	2%	2%	4%	5%	5%	5%
	8	5%	4%	3%	3%	4%	4%	6%	5%	5%	5%
	9	5%	6%	3%	4%	3%	3%	5%	5%	4%	4%
	10	12%	11%	10%	11%	12%	12%	12%	12%	10%	10%
	11	62%	62%	57%	57%	66%	66%	54%	54%	55%	55%
Revenu	1	4%	4%	9%	9%	5%	4%	9%	6%	0%	0%
	2	2%	2%	3%	3%	2%	2%	9%	10%	0%	0%
	3	2%	1%	4%	4%	2%	2%	2%	3%	0%	0%
	4	3%	3%	3%	3%	2%	2%	2%	2%	0%	0%
	5	2%	2%	4%	4%	2%	2%	3%	3%	0%	0%
	6	2%	2%	4%	4%	2%	3%	6%	6%	0%	0%
	7	2%	3%	3%	3%	4%	3%	4%	4%	0%	0%
	8	4%	4%	3%	2%	3%	3%	5%	6%	0%	0%
	9	3%	3%	3%	3%	3%	2%	5%	4%	0%	0%
	10	28%	26%	12%	13%	22%	22%	8%	9%	0%	0%
	11	48%	48%	53%	53%	56%	56%	48%	48%	99%	99%
Conversions	1	1%	1%	2%	2%	6%	7%	8%	8%	1%	1%
	2	1%	0%	1%	0%	12%	12%	11%	12%	1%	1%
	3	0%	1%	0%	0%	2%	2%	3%	2%	0%	1%
	4	1%	1%	1%	1%	2%	2%	2%	2%	1%	1%
	5	0%	0%	0%	0%	2%	1%	2%	2%	1%	1%
	6	2%	2%	1%	2%	2%	3%	2%	2%	2%	2%
	7	1%	2%	3%	3%	2%	2%	2%	3%	1%	2%
	8	1%	1%	2%	1%	5%	4%	2%	2%	2%	2%
	9	1%	2%	2%	3%	5%	4%	5%	5%	3%	2%
	10	40%	40%	35%	34%	14%	14%	17%	17%	38%	38%
	11	51%	51%	54%	54%	49%	50%	46%	45%	50%	50%
Visites	1	4%	10%	16%	15%	3%	3%	3%	3%	3%	3%
	2	1%	2%	2%	2%	1%	1%	2%	1%	2%	3%
	3	4%	3%	3%	3%	4%	3%	2%	2%	2%	2%
	4	5%	4%	4%	5%	4%	4%	5%	5%	4%	4%
	5	7%	6%	6%	6%	7%	7%	7%	7%	7%	6%
	6	10%	8%	8%	8%	6%	6%	9%	9%	9%	9%
	7	11%	10%	6%	7%	9%	8%	12%	12%	9%	9%
	8	9%	8%	6%	6%	11%	11%	9%	10%	10%	10%
	9	8%	7%	7%	7%	7%	8%	10%	10%	9%	9%
	10	41%	34%	31%	31%	49%	48%	42%	42%	45%	45%
	11	0%	8%	12%	12%	0%	0%	0%	0%	0%	0%

Tableau B.3 Variation de l'*avgCPC* pour chaque variation positive de position

	Variation de position	Graphe 1	Graphe 2	Graphe 3	Graphe 4	Graphe 5	Graphe 6	Graphe 7	Graphe 8	Graphe 9	Graphe 10
Profit	1	13.00%	8.00%	4.00%	6.00%	12.89%	3.56%	8.94%	13.00%	9.87%	9.75%
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-
	10	-	-	-	-	-	-	-	-	-	-
Revenu	1	13.00%	0.00%	1.11%	-	5.56%	0.00%	9.67%	9.29%	1.06%	-3.50%
	2	-	1.00%	-1.67%	-	-	-	-	-	1.78%	-
	3	-	9.00%	-	-	2.00%	9.00%	-4.00%	-	4.43%	-
	4	6.00%	-3.00%	0.00%	-	-	-	-	-	-3.71%	-
	5	-	-2.00%	6.00%	-	-	-	-	-	0.00%	-
	6	-	1.00%	-	-	-	-	-	-	-1.00%	-
	7	-	-	-	-	-	-	-8.00%	-	-	-
	8	-	13.00%	-	-	-	-	-	-	-1.00%	-
	9	-	-3.00%	-	-	-	-	-	-	-	-
	10	-	-3.67%	-	-	-	-	-	-	-5.50%	-
Conversions	1	0.21%	13.67%	5.90%	-5.50%	-1.00%	14.00%	16.50%	4.09%	7.06%	1.60%
	2	-	-	7.33%	-	0.50%	-5.00%	-3.50%	9.00%	0.00%	1.50%
	3	2.00%	-	-	-	2.00%	9.00%	-	-4.00%	-	-
	4	-	-	-	10.00%	0.00%	0.00%	0.00%	4.00%	2.00%	-
	5	-3.00%	-	-	-8.00%	-	-	-	-	0.00%	-
	6	-	-	-	-	-	-	-	10.00%	-	-
	7	-	3.00%	-	-	-	-	-	3.00%	8.00%	-
	8	-1.00%	-	-	-	-	-	-	-	9.00%	-
	9	-	-	-	-	-	-	-	-	14.00%	-
	10	-	-	-	-	-	-	-	-	8.00%	-
Visites	1	11.11%	13.50%	1.38%	0.64%	2.16%	-0.44%	11.05%	7.00%	9.03%	8.79%
	2	-	-	-	1.33%	-1.50%	-5.00%	3.67%	-	1.50%	-
	3	-	-	-	-5.50%	-	-	-6.00%	-	3.00%	-
	4	-	-	-	1.00%	-	-	7.00%	-	-	-
	5	-	-	-	-4.00%	-	-	4.00%	-	-	-
	6	-	-	-	1.33%	-	-	8.00%	-	-	-
	7	-	-	-	-7.00%	-	-	-7.00%	-	-	-
	8	-	-	-	8.00%	-	-	5.00%	-	-	-
	9	-	-	-	1.00%	-	-	12.33%	-	-	-
	10	-	-	-	-	5.00%	3.00%	-	-	-	-

Tableau B.4 Variation de l'*avgCPC* pour chaque variation négative de position

	Variation de position	Graphe 1	Graphe 2	Graphe 3	Graphe 4	Graphe 5	Graphe 6	Graphe 7	Graphe 8	Graphe 9	Graphe 10
Profit	-1	-14.60%	-10.67%	-8.00%	-13.00%	-10.46%	-2.54%	-10.07%	-11.00%	-9.05%	-10.60%
	-2	-	-	-	-	-	-	-	-	-	-
	-3	-	-	-	-	-	-	-	-	-	-
	-4	-	-	-	-	-	-	-	-	-	-
	-5	-	-	-	-	-	-	-	-	-	-
	-6	-	-	-	-	-	-	-	-	-	-
	-7	-	-	-	-	-	-	-	-	-	-
	-8	-	-	-	-	-	-	-	-	-	-
	-9	-	-	-	-	-	-	-	-	-	-
	-10	-	-	-	-	-	-	-	-	-	-
Revenu	-1	7.00%	-0.17%	2.17%	-14.00%	-8.25%	-3.88%	-8.40%	-10.25%	-1.13%	-
	-2	-	3.50%	-3.00%	10.00%	-3.00%	-4.00%	13.00%	-	-0.31%	-
	-3	-	-3.67%	3.00%	14.00%	0.00%	7.50%	-	-2.00%	3.67%	-
	-4	-2.00%	1.00%	-	-	-	-	-	-	-3.20%	-
	-5	-	-2.50%	-	-	-	-	-11.00%	-	8.50%	-
	-6	-	-	1.00%	-	-	-	-	-	2.00%	-
	-7	-	2.00%	-	-	-	-	-	-	-3.00%	-
	-8	-	-	-	-	-	-	-	-	-	-
	-9	-	-4.50%	-	-	-	-	-	-	0.00%	-
	-10	-	-16.00%	-	-	-	-	-	-	-	-
Conversions	-1	-3.63%	-18.00%	0.00%	-	-9.67%	-4.33%	-6.91%	-4.53%	-5.79%	-0.58%
	-2	-1.40%	-	-8.00%	-	3.00%	7.00%	9.00%	-6.00%	12.00%	4.00%
	-3	-1.00%	-	-	-	-	-	-	-5.50%	1.00%	-
	-4	-11.00%	-	-	-	-	-	-	-	-	-
	-5	-10.50%	-	-	-	-	-	-	4.50%	-	-
	-6	-	-	-6.50%	-	-	-	15.00%	-13.00%	-4.00%	-
	-7	-2.00%	-	-	-	-	-	-	-	1.00%	-
	-8	-	-	-	-	-	-	-	5.00%	-5.00%	-
	-9	-4.00%	-	-	-	-	-	-	-15.00%	1.50%	-
	-10	5.75%	-	-	-	-	-	-	-	-	-
Visites	-1	-5.67%	-10.86%	-5.77%	1.12%	-6.66%	-2.06%	-6.50%	-8.52%	-10.93%	-10.21%
	-2	-	-	-0.67%	-0.97%	6.00%	5.50%	-5.00%	8.00%	-0.50%	-
	-3	-	-	4.75%	0.36%	-2.57%	-4.14%	15.00%	3.00%	1.00%	-
	-4	-	-	-2.50%	-2.00%	1.20%	1.60%	-	-	1.00%	-
	-5	-1.00%	-	8.50%	3.27%	-6.00%	-10.50%	0.67%	-	-	-
	-6	-	-	4.00%	0.78%	2.63%	4.25%	13.00%	-	-	-
	-7	-	-	-3.60%	0.33%	1.40%	-5.20%	11.00%	-	-	-
	-8	-	-	-1.25%	-2.33%	-1.75%	-4.00%	0.67%	-	-	-
	-9	-	-	-1.75%	-	-1.00%	2.00%	-	-	-	-
	-10	-	-	-	-	-16.00%	-2.00%	-	-	-	-